

An investigation into the impact of retrieval practice strategies on science teachers and students

Penelope Robotham

A Research & Development Project

Submitted for the MSc Learning & Teaching 2020

DEPOSIT AND CONSULTATION OF THESIS

One copy of your dissertation will be deposited in the Department of Education Library via the WebLearn site where it is intended to be available for consultation by all Library users. In order to facilitate this, the following form should be completed which will be inserted in the library copy your dissertation.

Note that some graphs/tables may be removed in order to comply with copyright restrictions.

Surname	Robotham
First Name	Penelope
Faculty Board	Education
Title of Dissertation	An investigation into the impact of retrieval practice strategies on science teachers and students

Declaration by the candidate as author of the dissertation

1. I understand that I am the owner of this dissertation and that the copyright rests with me unless I specifically transfer it to another person.
2. I understand that the Department requires that I shall deposit a copy of my dissertation in the Department of Education Library via the WebLearn site where it shall be available for consultation, and that reproductions of it may be made for other Libraries so that it can be available to those who to consult it elsewhere. I understand that the Library, before allowing my dissertation to be consulted either in the original or in reproduced form, will require each person wishing to consult it to sign a declaration that he or she recognises that the copyright of this thesis belongs to me. I permit limited copying of my dissertation by individuals (no more than 5% or one chapter) for personal research use. No quotation from it and no information derived from it may be published without my prior written consent and I undertake to supply a current address to the Library so this consent can be sought.
3. I agree that my dissertation shall be available for consultation in accordance with paragraph 2 above.

Abstract

This project describes an investigation into the effect of retrieval practice on student performance in AO1 questions. Analysis of student performance in AO1 questions in chemistry showed their mean average to be lower than their performance in AO2 and AO3 questions, when compared to similar centres.

A review of the literature on retrieval practice shows how testing can improve student learning outcomes. Interviews with teachers at the start of the project revealed that knowledge of key facts was assumed to be taking place across the science department without explicit strategies to support learning. Interviews also revealed that teacher understanding and use of retrieval practice strategies was limited.

An intervention was designed to improve student performance in AO1 questions through the use of retrieval practice strategies. Weekly retrieval practice quizzes were developed using interleaved questions from different topics, focussing on AO1 questions as identified by the exam board.

Analysis of student data at the end of the intervention showed an increase in performance in not only AO1 questions, but also in AO2 and AO3 questions, compared to a previous cohort.

Feedback from students post intervention suggested that the introduction of regular testing also reduced their test anxiety.

Table of Contents

Introduction	7
Rationale for the project	7
Context of the project	9
Focus of the intervention	9
Impact of Covid19 on this work	10
Literature review	11
What is meant by learning?	11
How does learning take place?	14
Cognitive load theory model of learning.	15
Retrieval practice and testing effect	20
Types of retrieval practice	24
The frequency of testing	25
The role of feedback in the intervention	27
The effect of testing on student motivation	29
Conclusions	31
Methodology	33
Practitioner Research	33
The choices of research methodology based on research studies.	34
Research instruments	35
Obtaining teacher views on retrieval practice through interviews	44
The use of quantitative data analysis to evaluate the impact of the intervention on student performance	46
Obtaining student opinion about the intervention	47
Student questionnaires	48
Ethical considerations	49
The participants in the study	51
The format for obtaining teacher-researcher collaborators	52
The intervention	52
Feedback and marking of quizzes	55

Findings and discussion	59
Data analysis	60
RQ 1 - How do teachers and student currently make use of retrieval practice?	60
Knowledge base vs skills base	61
How do we, as teachers, know that learning has taken place?	64
Teacher's Prior knowledge about retrieval practice	65
RQ 2 - The impact of retrieval practice on student performance in AO1 questions.	66
Student views on the impact of the intervention	69
The impact of additional testing on student anxiety?	71
Independent use of retrieval practice	73
Conclusions	76
Validity of conclusions	77
What was the impact of an intervention on retrieval practice on student performance outcomes in AO1 questions?	78
Validity of conclusions	80
Student views of the intervention	80
Unplanned outcomes	83
The next steps	83
References	85
Appendices	98

List of figures

	Page
Figure 1 Exam question from 2018 AQA Chemistry 7405 Paper 1	13
Figure 2 Kolb's learning cycle (1974)	14
Figure 3 The three stages of memory as described by Melton (1963)	15
Figure 4 Caviglioli's adaptation of Willingham's simple memory model (2018)	16
Figure 5 An example of a typical quiz	53
Figure 6 An example of a student work with self-corrected their answers	56
Figure 7 A student record sheet	57
Figure 8 Comparison of student performance in AO1 questions pre and post intervention	67
Figure 9 Comparison of student performance in AO2 questions pre and post intervention	68
Figure 10 Comparison of student performance in AO3 questions pre and post intervention	68
Figure 11 Student responses to exam anxiety	72
Figure 12 The impact of additional testing on student anxiety	72
Figure 13 Student independent use of retrieval practice questions	74
Figure 14 How students space retrieval practice questions	75

List of tables

	Page
Table 1 Assessment objective (AO) weightings for AQA A' level chemistry	6
Table 2 Methodology choices	42
Table 3 Average score of control group vs intervention group	69

Introduction

Rationale for the project

This dissertation describes a collaborative practitioner research project which focusses on the impact of retrieval practice on student performance in AO1 questions. The work includes a literature review looking at current research on retrieval practice and its use in education. The methodology is an easily reproducible process designed to increase the use of retrieval practice in the classroom by student and teachers.

As a classroom teacher I have experienced the frustration that comes from teaching an idea or concept to students, that they seem to 'get' at the time, only to find that they later can't remember it. Analysis of my A' level student's performance in their chemistry examinations highlighted that, although their overall performance was much higher than the national average, their performance in their AO1 (knowledge and recall) questions was weaker than their performance in AO2 (application) and AO3 (analysis and evaluation) questions, compared to similar centres. This was surprising to me, as the cohort of students were from China where their education, prior to studying in the UK, primarily focussed on the factual recall of knowledge, which is typical of AO1 style questions.

The project aims were to investigate the effect of retrieval practice strategies on student performance, focussing strategies on AO1 questions. An analysis of current retrieval practice used within the science department at the start of the project was conducted through interviews with the college science teachers. An intervention strategy was developed and trialled and the impact of the project evaluated by analysis of student responses to AO1 questions by comparison of the intervention group student performances to data from the previous cohort, who did not participate in the study. Collaboration between chemistry teachers in my college and with teachers across the science department is also discussed.

The Office of Qualifications and Examinations Regulation (Ofqual) regulates qualifications in England, setting assessment objectives (AOs) across all UK A' level chemistry specifications and exam boards. In A' level chemistry the three AOs assess a student's ability to:

- Demonstrate their knowledge and understanding of scientific ideas, processes, techniques, and procedures. This is known as AO1.
- Apply their knowledge and understanding of scientific ideas, processes, techniques and procedures, in theoretical and practical contexts and when handling qualitative and quantitative data. These types of questions are categorised as AO2.
- Analyse, interpret and evaluate scientific information, ideas and evidence, including in relation to issues, to make judgements and reach conclusions and to develop and refine practical design and procedures. These skills are categorised as AO3.

Within my College students follow the AQA A' level chemistry specification, and the weighting for each AO, for this specification are showing in table 1.

Table 1 Assessment objective (AO) weightings for AQA 7405 GCE A' level chemistry

Assessment Objective (AO)	Overall weighting %
AO1	30
AO2	45
AO3	25

Context of the project – school background

The project was carried out in a UK independent International college, which specialises in STEM education. The majority of the students are from overseas and residential, with 98% of the students classified as EAL learners. The College is selective and requires students to pass assessments in maths and English prior to being offered a place and in 2019-2020 had 56 students on roll. The majority of students follow a two year A' level programme with some following a 3 year programme with intensive English language tuition. The College achieved over 80 % A*- B grade at A' level in 2019, with 64% of grades achieved A* - A .

Focus of the innovation

Within the College there is a constant drive, and expectation, for nothing short of excellence in attainment, from a mixed ability range of international students. The students arrive at the College with very different experiences of education from across the world. Therefore, the necessity to continually review and improve teaching techniques to give all students the very best chance of success, in chemistry in particular, was very much a motivator for this project from its inception.

Throughout my teaching career, the majority of testing that I have carried out has been summative, end of topic assessments, with the primary purpose of providing evidence of learning *outcomes*. This type of practice mirrors the most common use of testing as reported by Kibble (2017). Prior to this work my use of tests to enhance learning is has been limited.

Retrieval practice, also known as the testing effect, is the observation that retrieving information from memory strengthens both the durability of that information within the long-term memory, and the ability to retrieve it in the future. Such practice was described by Binks (2018) as “one of the most robust phenomena in cognitive psychology”, yet she observed that teachers and learners do not draw upon it to support learning. The potential impact of retrieval practice, as a strategy to improve my student performances, is explored within this work.

Impact of Covid19 on this work

The initial plan for the intervention was to develop a set of retrieval practice resources that could be used with Year 2 A' level chemistry students. Their performance in A' level external examination in the Summer series 2020 was then to be analysed and the impact of retrieval practice assessed by comparing their performance in AO1 questions to those of a control group of previous students from the College. Data from the A' level examinations would be both summative and 'high-stakes' in that the outcomes would enable the students to progress to University.

The Covid19 outbreak led a very short notice national lockdown and the cancellation of all external examinations in March 2020, therefore, analysis of data from the A' level papers was not feasible. As a result the project was amended and the impact of the intervention was measured against performance of Year 2 students in their mock examinations held just before the lockdown was implemented. It is appreciated that this data was not as high stakes for the students, and that some may not have put the same effort and preparation into these mock examinations as they would into a final examination and as such this data is not as robust as the data that I had intended to analyse. However, question by question analysis to compare the performance of students post intervention to the previous cohort using identical examination questions was now possible as part of the analysis of the impact of the intervention. The previous cohort had been taught by the same teacher and conclusions drawn from the data are framed with these considerations in mind.

Literature Review

This literature review begins with a review of academic definitions of learning which are considered in the context of A' level students. As factors that affect how students learn are complex and multi-faceted, I chose to focus my literature review on research concerning how students learn in terms of cognitive load theory, CLT, which was a basis of my Part 2 research project and this study follows on from work conducted during my Part 2 study.

Through my Part 2 work, the potential of retrieval practice was identified as being suited to my aims for this project, as a potential tool for improving student performance. I therefore include here a review of the literature around retrieval practice, considering the different types of practice reported, to determine which might be most suitable for this study. Finally, I consider the literature around the role of feedback to elicit an intervention strategy to best suit the needs of the college and my students.

As this project specifically concerns the impact of retrieval practice on student performance in AO1 questions, this review discusses only the strategies that I used post-teaching to develop my students ability to retrieve information that has already been taught. I appreciate that there are many hidden variables which may affect a student's ability to learn, such as their prior learning and their misconceptions, however, due to the timescale and nature of this project, these are not discussed here.

What is meant by 'learning'?

I trained and qualified to teach in the late 1990s, when there was a focus on differentiation and learning styles, and e-learning was not yet a concept. Throughout my ITT and teaching career since, the focus of INSET has been concerned with strategies to get information **into** students' heads. Little emphasis has been placed on how to get information 'out' again or how learning might be measured, other than the use of the end of key stage high stakes tests, linked to performance

league tables. Through INSET focussed on perfecting the instructional regime, the implication has been that 'learning' (as measured by assessment regimes) would follow. The lack of a universal understanding of what is meant by 'learning' began to concern me more as I developed as a teacher. This concern mirrors that of De Houwer et al (2013) who noted that researchers were rarely explicit about what they meant by learning. Whilst commonly recognised throughout the literature that learning is a continuous process, Kelly (2002) stated that learning "is a very individual, complex and, to some degree, an indescribable process" which describes the difficulty of trying to define 'learning'.

The word 'learning' is used routinely in discussions about teaching. Educational researchers agree that learning is much deeper than memorization and information recall, describing it as a process that is active, building on prior knowledge and requiring cognitive engagement. For example, Dewey, Piaget and Vygotsky (in Pardjono 2016), have all described the active process of learning, with learners building knowledge through observations and interactions with the world around them. Traditionally, studies around learning have focussed primarily on early years learning through childhood and adolescence with few exploring on the age profile of the students concerned in this work (16-18 years).

In studies with adult learners, who are closer in age to my students, to explore their understanding of learning, Säljö, (1979) reported them describing it in terms of 'acquiring' and 'memorising' information, with the 'active' process of learning not recognised. This perception of 'learning' by older students relates to AO1 type questions, where learners are required to demonstrate their knowledge and understanding of scientific ideas.

Whilst the size and nature of content material varies by subject, many subjects have a body of factual knowledge that needs to be committed to long-term memory in order for students to access problem-solving, AO2 and AO3 type questions, (Glaser & Chi, 1988; Hattie & Timperley, 2007).

Studies reports that the transfer of knowledge or skills to a novel problem requires both knowledge of the problem's context and a deep understanding of the problem's underlying structure, (Bransford et al., 2000; Pellegrino & Hilton, 2012). Therefore, in agreement with Pellegrino & Hilton, (2012) and Day & Goldstone, (2012), my role as a teacher should ensure that my students have sufficient background knowledge so that they can demonstrate higher order thinking skills.

Question 4.5 from the AQA 2018 chemistry paper, figure 1, illustrates how recall of key ideas is necessary in order to be able to answer the question.

The figure originally presented here cannot be made freely available via ORA because of copyright. The figure was sourced at AQA, (2015) AS and A Level Chemistry Specification, [Online]. Available from: <https://filestore.aqa.org.uk/resources/chemistry/specifications/AQA-7404-7405-SP-2015.PDF>

In this example, students must be able to recall key ideas such as, the chemical test for BaSO_4 and the solubility of this compound to apply this knowledge to the question. Therefore, improving student performance in AO1 questions, may also improve their performance in AO2 and AO3 questions, supporting the rationale for this project.

Definitions of learning within the academic literature, describe a measure of learning as the ability of students to be able to demonstrate independent and critical thinking and to transfer knowledge to new and different contexts (Barnett & Ceci, 2002). Because learning is so complex and a simple definition does not exist, for learning to take place in the context of this project, I have focussed this work on the development of strategies that would enable a student to recall or pull information back from their memories when needed. It is appreciated that learning is far more complex than this, such as the consideration of student misconceptions, but due to the timescale of the project these are beyond the scope of this current work.

How does learning take place?

There are a wide range of theories that attempt to explain how people learn. Such theories often differ from each other depending on the type of learning that they describe and learning theories associated with children and adolescents differ from theories associated with adult learning. The students in the context of this project lie in the gap between adolescence and adult learners, where little research has been reported. Academics have attempted to map and explain the learning process and some common themes appear in the literature. Generally, it has been reported that learning takes place in repetitive cycles, such as the Kolb's learning cycle (1984, p 21), where learning is viewed as an integrated process (figure 2). In this and other examples, the iterative process of learning is illustrated, with learning evolving as learners develop and use early knowledge for later understanding.

The figure originally presented here cannot be made freely available via ORA because of copyright. The figure was sourced at Kolb, D. A., (1984) *Experiential Learning: experience as the source of learning and development*. Englewood Cliffs, Prentice Hall.

Many learning models outline the stages of learning in terms of 'planning', 'doing', 'reviewing' and 'learning' and such cycles have provided a foundation for my thinking about how to support my student's learning in the classroom in this work. Through considering the use of retrieval practice in this project, I will be aiming to develop the use of the review stage of these learning cycles.

Cognitive load theory model of learning.

A model of learning that I explored in my Part 2 work, in considering how to improve student outcomes, was cognitive load theory (CLT), which was described by William (2017) as ‘the single most important thing for teachers to know’. CLT emerged from the work of educational psychologist John Sweller in the 1980s, (Sweller et al. 1988) and is based on a number of widely accepted ideas about how human brains process and store information (Gerjets et al. 2009, p. 44).

Commonly accepted models concerning human memory are built around a three part model, such as that proposed by Atkinson and Shrifin (1968), based on the work of Melton (1963). This model is summarised in figure 3, and is based on the concept of sensory inputs being processed in the short-term (working memory) before being transferred into the long-term memory.

The figure originally presented here cannot be made freely available via ORA because of copyright. The figure was sourced at Melton, A. W. (1963) Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*. (2) p 1-21

Figure 3 – The three stages of memory as described by Melton (1963)

Willingham (2009), proposed a similar three part memory model which shows the relationship between the working memory and long term memory systems. Learning is represented in this model by the transfer of information into and out of the long term memory as remembering. This model was adapted by Caviglioli (in Lovell, 2018), figure 4, and shows the relationship between external stimuli in our environment and the transfer of information received from the working memory (short term memory), into the long term memory.

The figure originally presented here cannot be made freely available via ORA because of copyright. The figure was sourced at Lovell, O. (2018). Sweller's Cognitive Load Theory in Action (In Action Series) Published by John Catt

Figure 4 – Caviglioli's adaptation of Willingham's simple memory model (2018)

In such multi-store memory models, information is described as being stored in the long-term memory in the form of schemas (Atkinson & Shiffrin, 1968; Tulving, 1972) and that the processing new information results in 'cognitive load' on the working memory which can affect learning outcomes (Anderson, 1977; Atkinson & Shiffrin, 1968; Baddeley, 1983). Retrieval and rehearsal of information has been reported to strengthen schemas and therefore the storage of knowledge (Atkinson & Shiffrin, 1968), thus fitting with the aims of this project and the use of retrieval practice as a method to improve student recall.

Cognitive load research aims to develop instructional techniques and recommendations that fit within the characteristics of working memory, in order to maximise learning. Clark et al. (2012) have described the working memory as 'the limited mental "space" in which we think' (2012, p. 8). In the context of this work, this is the memory system that a student would use whilst processing or answering an exam question. If students do not have fluency in their knowledge and are having to use some of their working memory to retrieve it, then they have less available to use to work through the question being asked. Willingham (2009), has said it is "absolutely crucial" that students store subject-specific knowledge in their long-term memories to allow them to develop "the thinking skills that are, for many teachers, the real goal".

Although the actual amount of information that an average person can hold in their working memory varies across the literature, there is an agreement that it is limited, (Cowan, 2001; Barrett et al, 2004). These findings have implications in the classroom, suggesting that if my students do not practise what they have learnt, then the knowledge and content that has been delivered will not be retained. Peterson and Peterson (1959) demonstrated that almost all information stored in the short-term memory is lost within 18-30 seconds if not rehearsed. The implication of such findings is that without the development of strategies to move content into the long-term memory the majority of the lesson content will be 'forgotten' before the next learning episode.

The process by which learners recall information from their long-term memory was defined by Sweller, et al (1998, p 256) as mastery. Cognitive scientists propose that mastery reduces the burden on working memory; that if information is rehearsed and retained beyond the working memory this will bypass the limits of the working memory, (Sweller, et al; 1998, p. 255; Laberge & Samuels; 1974). In the context of this work, developing practices for students to move information into their long-term memory would enable a student to achieve mastery. Without having to stop to consider each stage of the process, students would have a larger capacity in their working memory to process questions with the aim of improving examination performance. Therefore, the CLT model of learning fits in with the aims of this dissertation through developing a tool for long-term memory and automaticity of knowledge, with the goal of improving student performance in AO1 questions.

Whilst there are many positive outcomes is the use of CLT within the literature that report strong and lasting influence on learners in many educational situations there is also considerable debate over many of its principles and there are a number of criticisms of the model that I have considered. A criticism of CLT research is that many of the findings evolved from lab-based trials, with a control group typically learning using conventional techniques and a treatment group using cognitive load techniques.

Test performance of participants is taken as an indirect measure of cognitive load, with high results on post-tests considered to indicate that cognitive load was successfully managed, (Stull & Mayer 2007; de Jong, 2010, p 123-125). The difficulty of applying these findings to general classroom teaching environments has been raised as a criticism of the findings, however, CLT has been more recently applied to educational settings such as the work of Cooper & Sweller (1987) with high-school maths students.

Another criticism of CLT is that researchers are unable to standardise a method of measurement as to what constitutes cognitive load. The majority of CLT studies do not attempt to directly measure cognitive load, but rather measure the effectiveness of instructional techniques designed to consider the limitations of working memory. Participants of studies are typically asked to rate their cognitive load on a scale and such methods are highly subjective and vary from participant to participant. Studies which have attempted to directly measure the cognitive load imposed by different techniques, are not consistently reliable (e.g. de Jong 2010; Paas, Renkl & Sweller 2003) and it is the lack of methodical rigour of some studies that is a cause for concern by some (Brünken, Plass & Leutner 2003; de Jong 2010; Moreno 2010; Schnotz & Kürschner 2007).

The lack of empirical indicators to distinguish between and measure the different types of load is also criticised by many (de Jong 2010; Gerjets, Scheiter & Cierniak 2009; Schnotz & Kürschner 2007). Some attempts have been made to address this criticism (DeLeeuw & Mayer 2008; Leppink et al. 2014) and key supporters of CLT acknowledge the need to identify a reliable means of directly measuring load, in order to develop a more empirical basis to support the theory (Paas, Renkl & Sweller 2003, p. 4; Paas, Tuovinen, Tabbers & van Gerven 2003, p. 64).

The majority of studies on cognitive load do not consider how individual differences between learners might impact upon cognitive load, or how other factors besides cognitive load might influence the effectiveness of learning.

Moreno (2010) noted that cognitive load theory does not consider, for example, how factors such as a learner's motivation might influence the effectiveness of learning. There is a rich body of literature concerning factors which affect student motivation and its impact on learning.

Generally, the students in my class are very highly motivated, aspiring for places at Russell Group universities. Therefore, I acknowledge that such factors are important but less significant in the context of this study and are therefore not discussed here.

Despite concerns regarding aspect of cognitive load theory, it is supported by a large number of RCTs, and the body of research in support of cognitive load suggests that it is worth considering in terms of practice in the classroom and my current teaching. Meta-analysis has found that the effect of the size of lab studies is comparable to those carried out both in primary and secondary school classrooms, (Adesope et al. 2017). Therefore, by supporting my students to move knowledge into their long-term memory and developing techniques so that they practise this knowledge so that it becomes 'automated' I hope that consideration of this model of learning may have the potential to improve student performance in AO1 questions.

Cognitive load theory states that the capacity for storing memories is unlimited, but that the ability to access a given memory typically declines over time. The theory proposes that regular retrieval is required to enable students to be able to access such memories when needed. Many cognitive load theory studies focus on how to reduce factors that overload the working memory.

Little has been reported concerning how the theory may be applied to moving information into the long-term memory and the impact on student performance as a result of the working memory being freed up to tackle examination questions, which is the focus of this project. Therefore I will now consider how retrieval practice may be a strategy to develop student performance.

Retrieval practice and testing effect

There is a significant body of evidence that formative assessment is a powerful tool for improving outcomes for learners. A number of reviews of the research conducted have analysed several thousand studies and have reported the impact of classroom assessment practices on students' performance (Black & Wiliam, 1998; Hattie & Timperley, 2007). Although some of these reviews report the negative effects of some assessment practices, the majority show that, where used appropriately, formative assessment enhances student achievement (Wiliam, 2010). Retrieval practice is one form of formative assessment that was not well developed throughout my department at the start of the project, with the aim of this work to assess the impact of using it on student performance.

Throughout the research into memory there is a consistent finding that long-term retention of information by students is improved through the practice of 'testing' when compared to other study methods, (Einstein et al., 2012; Karpicke & Roediger, 2008; Roediger & Karpicke, 2006a). Evidence for the effect of testing has been reported across different types of learners, including primary aged children, (Jaeger et al., 2015; McDaniel et al., 2013; and Rohrer et al., 2010) college students (Carpenter & Pashler, 2009; Dobson & Linderholm, 2015; Jacoby et al., 2010; Roediger & Karpicke, 2006b), and professional trainees such as medical students, (Baghdady et al., 2014; Larsen et al., 2009).

The terms 'retrieval practice' or 'the testing effect', are used to describe the act of recalling learned information from memory, with no or little support. Several quasi-experimental and correlational studies have reported that testing or quizzing can enhance performance for students, (Angus & Watson, 2009; Daniel & Broida, 2004; Kibble, 2017). Dunlosky et al (2013) rated testing as one of only two learning techniques described across a body of literature as having 'high utility' for classroom practice.

There is a significant body of literature that has said that the process of taking part in a test is not a neutral event. Studies have demonstrated the potential pedagogical benefits of testing (Roediger & Karpicke, 2006b) and in recent years, cognitive psychologists have been comparing retrieval practice with other methods of studying, including the use of study guides, and re-reading texts. They have reported that long-term learning is significantly enhanced by retrieval practice compared to other methods of studying, (Roediger & Butler, 2011). Significant improvement for students on end of course examinations has been reported with students who were regularly quizzed scoring *a full grade higher on the material from the quizzes* than on material which they had not been tested on (McDaniel et al; 2011). Such findings have been reported for both online testing and for tests carried out in classroom environments. In addition, findings have been replicated in studies with both college students (McDaniel et al. 2013) and middle-school students (McDaniel et al, 2011 and Roediger & Butler 2011).

It has been proposed that when students have to actively recall information previously taught, that the process develops their abilities to recall the information they have learned, that the act of retrieving answers to questions during the testing can also develop and enhance the memory for the tested information, (Carpenter and DeLosh, 2005; Hogan and Kintash, 1971; Roediger and Karpicke, 2006a).

These findings suggest that there is potential in using testing a technique to develop learning in my students as well as in its more traditional form as a method of testing learning. However, on reviewing the methodology of such studies many demonstrate the effectiveness of testing by using identical questions in the final examination as those used in practice tests. In preparing my students for externally set A' level examinations, using the same questions in the testing process as in the final assessment is clearly not possible.

The limitations of retrieval practice studies is also supported by other critics. Many of the reported findings have been in cases that have focused on word lists and paired associates, for example Dunlosky et al (2013, p 32) leading questions to be asked about whether this strategy improves performance in more complex tasks. Van Gog and Sweller (2015, p247) argued that

'the testing effect decreases as the complexity of learning materials increases... the effect may even disappear when the complexity of the learning material is very high'.

Rohrer et al (2010) also reported that the

'benefits of retrieval practice have yet to be demonstrated for mathematical tasks other than fact learning'.

Despite the uncertainty about the range of materials for which retrieval practice is useful, as AO1 questions require learners to demonstrate their knowledge and understanding of scientific ideas and processes, retrieval practice been shown to have potential for these types of questions.

Explanations as to why retrieval practice has a positive effect on student performance in tests, commonly suggest that the process of testing promotes an active search for answers by a student and that this process in itself enhances and reinforces retrieval cues used by a student, Carpenter et al, (2009) and Rowland, (2014). Repeated retrieval practice attempts have been suggested to result in strengthening memory making future attempts at retrieval even more effective, Dunlosky et al., (2013), Karpicke and Blunt, (2011). Butler (2010) proposed that where testing uses different questions to elicit the same knowledge, that students develop multiple retrieval routes to that knowledge and that this increases the likelihood of successful future retrieval. If students are more able to successfully recall information, then this should increase their performance in AO1 questions.

In her comprehensive review about the effect of retrieval practice, Binks (2018) concluded that the evidence for the direct and indirect benefits of testing to learning and transfer was so strong and consistent that “educators would be remiss in not utilizing testing as a pedagogical strategy”. Bink’s concluded that the body of evidence in the literature shows that testing can be used not only to evaluate learning, for example at the end of the topic, but also to promote learning.

Whilst there are a significant number of cases strongly supporting retrieval practice and its potential in the classroom, not all of the effects reported may be directly linked to this practice. The knowledge that regular testing is being carried out may influence a student’s study habits as described by Roediger & Karpicke, (2006a). If testing is conducted frequently, students may be more likely to study more, therefore in feedback from students following the intervention, their study habits will be evaluated and considered when making conclusions.

Despite the number of studies showing the effectiveness of testing on learning, and perhaps because testing is often viewed as a method of assessment, rather than a method of learning, retrieval practice is not yet embedded in classroom practice. Yet retrieval practice appears to have potential in this project as a strategy to increase student performance in AO1 questions.

Cognitive scientists propose that every time that information is retrieved, it changes that original memory, making it stronger and more likely to be recalled at a later time. Bjork (1994) described such practice in terms of *‘using your memory, shapes your memory’*. It is claimed that the process of retrieval cements information in the long-term memory, which should enable it to become easier to recall in the future. This explanation of the benefits of repeating to learn matches my own experiences repeating times table so become automated and therefore in terms of my own learning seems an appropriate method to explore in more detail in my own classroom.

A number of studies have also put forward evidence to suggest that testing improves transfer of knowledge to new contexts.

'Retrieval practice does not simply enhance retention of the individual items retrieved during the initial test: taking a test can also produce retrieval-induced facilitation – a phenomenon that shows testing also improves retention of contested but related material.' Roediger & Karpicke 2006a

In my context this would suggest that, not only may retrieval practice improve student's performance in AO1 questions, but improvements may also be observed in AO2 and AO3 questions where students are required to transfer and apply their knowledge. Therefore from the large body of supporting evidence for retrieval practice, I believe that it merits further exploration of this practice in this project.

Types of retrieval practice

Retrieval practice focuses on recalling information from memory as a learning tool and it has potential to become part of my classroom practice to support learning. This section reviews the literature surrounding the types of retrieval practice techniques reported to determine an appropriate strategy to implement and embed in the classroom.

In her review of classroom retrieval practices, Jones (2019) summarised a number of techniques, including low stakes quizzing and flashcards as examples of retrieval practice. In addition the use of writing prompts have also been reported to be "powerful tools to improve learning" (Chan, 2009).

Although the use of flashcards as a retrieval method has been reported widely in the literature Karpicke et al's (2009) research suggests that misuse the technique results in the degree of retrieval being significantly reduced. Methods that require students to actively try to retrieve the answer themselves have been reported to have a greater benefit on student's learning. Therefore, although I regularly use flashcard techniques as revision tools in my classroom, for example Quizlet exercises, I chose not to build upon this style of retrieval practice in this intervention.

Reported practices throughout the literature show that when the purpose of testing is to promote rather than assess learning, as is the goal of this project, tests that require learners to produce answers to questions are superior to multiple-choice or other types of recognition tests (Butler and Roediger, 2007, Kang et al., 2007, McDaniel et al., 2007, McDaniel et al., 2007). Studies also report that retrieval is more difficult if the testing conditions require students to recall or produce of information compared to recognising answers such as in multiple choice tests (Jaeger et al., 2015, Rowland, 2014) and that the harder the retrieval process the deeper the learning (Rowland, 2014)

In her review of studies on retrieval practice, Binks (2018) also concluded that for it to be most effective, students should be required to produce their own answers, rather than recognising answers from multiple choice style questions. Therefore in designing an intervention strategy around retrieval practice, the evidence suggests that it would be advisable to use open response questions rather than multiple choice questions. As the majority of my students are EAL learners, following the findings of Jaeger (2015) and Rowland (2014), short answer questions were considered to be most appropriate for the nature of this study, with an additional benefit of helping to develop my students skills in writing extended answers. By using short answer style questions, it would remove the possibility of students 'guessing' the correct answer from those provided.

The Frequency of testing

With the goal of designing an intervention that supports my students in recalling material taught throughout the two year A' level programme and to increase student performance in AO1 questions, the frequency of testing different items was carefully considered and discussed in collaboration with my other teaching colleague in designing the materials we trialled. As Pyc & Rawson (2009) reported that the more challenging the retrieval task, the better the learning here I consider evidence from studies investigating the frequency of testing to support the design of the intervention.

Interleaved practice involves testing students' recall across different topics. In Binks' review of retrieval practice (2018) she concluded that there was clear evidence regarding how and when testing should be conducted, with studies showing that interleaved, or spaced practice, results in students having significantly better long-term retention and transfer of skills. Such benefits of spaced practice on long-term retention, have been demonstrated across subjects, types of students and time scales.

Other studies have also shown that interleaving practice is better at developing a student's long-term retention, as measured when participants take a test on the material in delayed tests, taken after a period of time has passed since the intervention (Simon & Bjork, 2001). The size of the long-term advantage of interleaved practice has been reported to be extremely significant (Rohrer, Taylor & Sholar, 2010). In addition, studies have also suggested that because new learning depends on prior learning, spaced learning may also enhance the transfer of knowledge and provide a foundation for subsequent new learning, (Kornell & Bjork 2007). In comparisons of interleaved practice compared to mass practice, participants who had practised under interleaved conditions performed far better than the blocked-practice participants, "who appeared, when tested under a random schedule, to have learned virtually nothing", (Shea & Morgan, 1979).

Several theories as to why the practice of interleaving enhances long-term retention and transfer have been proposed. Switching between different questions and topics has been proposed to develop a learners ability to identify similarities and differences between them which may enable the student to encode ideas using higher-order representations. An alternative explanation suggests that interleaving forces learners to 'reload memories'. Repeated reloading of memories is thought to improve learning so that the knowledge or skill may be accessed when needed at a later time, (Bjork & Bjork, 1992).

The evidence throughout the literature, suggests that the most efficient and effective testing regimen appears to require one correct retrieval of information during initial learning followed by three or four subsequent tests (Rawson & Dunlosky, 2012). Reported evidence indicates that tests should be spaced at expanding versus equal intervals, over days or weeks, to maximize retention (Kang et al. 2014) and such factors were taken into consideration in the design of the intervention.

The role of feedback in the intervention

This section reviews the literature regarding feedback to determine a rationale for its use in the intervention.

A substantial number of reviews and meta-analyses explore the effects of feedback and whilst there are differences in the reported benefits of feedback on improving student progress, the consensus is that feedback given to the student, or teacher, about the student's performance relative to learning goals or outcomes, leads to improvement in students' learning, (EFF, 2018). However, the effect on learning is wide-ranging and some studies show that feedback may have negative effects. It is therefore important to understand the potential benefits and the possible limitations of feedback when developing the intervention strategy.

It is commonly agreed throughout the literature that, in order to inform their learning, students need to receive feedback on their work and performance to know the next steps to take. The provision of high-quality feedback can lead to an average of eight additional months' progress over the course of a year, with feedback taking a range of different forms, including written, oral, and peer feedback (EFF, 2018). The importance of effective feedback strategies is further enhanced by Hattie's (2008) wide-ranging review of some 800 meta-analyses of the influences on pupil learning, leading him to conclude that the most powerful single influence enhancing achievement is feedback. Hattie reported that the effect size of feedback was 1.24 and concluded that feedback has more impact on learning than any other general factor.

In her comprehensive review of the literature concerning the effect of testing on learning Binks (2018) concluded that the evidence showed that in order for testing and retrieval practice to be most effective, feedback should be provided for all questions set for students.

Studies have shown that feedback enhances learning outcomes (Butler & Roediger, 2008) and that optimal feedback should indicate not only the correct answer, but also incorrect answers, alongside an explanation as to why answers are correct or incorrect, (Butler et al 2013). Student learning can be increased when students receive regular, specific, explanatory, and timely feedback on their work and the feedback teachers offer can be most effective when it provides students with specific information about their current state of knowledge and performance as related to learning goals (Hattie and Timperley 2007). Sharing with students their level of performance so they can direct and evaluate actions and efforts has been reported to be beneficial to learning (Locke & Latham 1992). Such feedback enables students to track their performance in relation to target grades so that they can adjust their effort or learning strategies as needed.

By using self-testing or practice assessments on material students are to learn, Dunlosky, et al (2013), also reported that assessment itself can have a positive impact on learning. Assessments would then move to being used formatively rather than in the summative manner more commonly seen and reported by Wiliam (2009) and Leahy et al. (2005).

Thus, in developing an intervention strategy around retrieval practice, the work within the literature highlights the importance of including feedback in the development of the intervention. In developing the intervention, based on the literature review, I felt that it was important that my students were told when an answer was wrong, or if they had not covered all of the relevant information, after their attempts to recall information.

However, while the *average* impact on learning is high, studies show that in some cases feedback can have negative effects on attainment. The effectiveness of feedback has been shown to be dependent on several factors, including the ability and motivation of the student. As every student in the intervention group is very able and highly motivated, this concern is of less significance in this intervention that it might be in an alternative setting.

Studies have also shown that the tone and targeting of feedback affects student motivation and that students tend to respond better if feedback addresses significant aspects of their work and understanding, (Brookhart, 2008; Minstrell, 2001). These findings were incorporated into the feedback strategies developed as part of the intervention.

The effect of testing on student motivation

One of the benefits of retrieval practice that has been reported in the literature, is the increase in student motivation, Binks (2018). Students' motivation to learn may be described in terms of goal theory, with two broad types of goals identified within the literature. Mastery goals have been described as those that are focussed on increasing competence and improving performance and gaining mastery of the subject, whereas performance goals are focused on outperforming other students (Poortvliet et al., 2007). A student who is driven by performance goals may avoid challenges if they are concerned about not performing as well as other students. In my classroom situation, where I am striving for every student to achieve their full potential, I aim to encourage students to adopt mastery goals, to understand the material as best as they can, rather than be concerned about the performance of others in the class. To foster this type of goal within the classroom the intervention was designed to enable student evaluations of their work to be done individually and privately, thus avoiding social comparisons.

Test-taking motivation has reported to vary most in low-stakes situations (Barry et al, 2010) with students displaying greater variation in the amount of effort they apply in such situations (Theilk et al, 2009).

In contrast, in high-stakes assessments, such as A' level examinations, Cole et al (2008) suggest that motivation is considered less of an indicator of outcome because the level of motivation is generally higher in such testing situations. Low-stakes testing, where the outcome does not have any impact on a student's grade, have been associated with a lack of motivation by students to perform well in them. As a consequence, it has been suggested that students achieve below their best level of academic achievement (Eklöf, 2006; Eklöf & Nyroos, 2013; Liu et al, 2012; Theik et al. 2009).

Students, who are identified as being well-motivated, have been shown to perform better in achievement situations, expend more effort learning new tasks, use more efficient self-regulatory strategies and persist longer on difficult tasks (Eklöf, 2010). Testing, however, may only motivate some students and therefore the use of tests may increase the gap between higher and lower achieving students (Knehta, 2017). Similar conclusions have also been made by Harlen & Deakin Crick (2002) in their meta-analysis of impact of summative assessment and tests on students' motivation for learning. Whilst mindful of these reported concerns regarding testing, the students in the intervention group are all well motivated and complete work on time, and display high levels of engagement with tasks. It is very rare that any work is late or incomplete and where this has occurred, this has been rectified through one-to-one conversations with individual students. Therefore, the negative impact of testing reported in some studies is considered to be less likely with this particular cohort of students, but was taken into consideration when the intervention was introduced to students, and followed up in student questionnaires.

Conclusions

To guide my review of the relevant literature for this project, I used four questions. Firstly, I wished to define what is meant by 'learning' in the context of an A' level student. As de Houwer (2013) noted, researchers are rarely explicit about what they mean by the term "learning" and many textbooks on learning do not always contain a definition of its subject matter. The broad and abstract concept of learning has proven difficult to define throughout the literature. Whilst many academic definitions of learning discuss a long term change in behaviour, the most appropriate definition of learning, for my context is that of Barnett & Ceci, (2002) who describe it in terms of a student's ability to recall and then transfer their knowledge to new and different contexts, as I am only with my students for the two years of their A' level programme and measuring long term changes is therefore restricted. Therefore, to measure the impact of this intervention a comparison of student performance on AO1 questions compared to a control group where no intervention had taken place was taken as a measure.

Following on from my Part 2 research, I considered the literature concerning cognitive load theory as a model for how learning takes place. Cognitive load theory considers that during learning information is stored in the working memory until it has been processed sufficiently to pass into the long-term memory. It provides considerations for practices that would enable students to move information into their long term memory and deepen their learning. These considerations have been used to develop my intervention. I then researched studies into retrieval practice, as a strategy for improving student performance. Research reported many benefits of the use of retrieval practice and testing that may have a significant impact on student performance. Students employing such techniques are reported to perform significantly better when compared to other study methods. Such studies support the aims of this project to explore using retrieval practice as a strategy to improve student performance in AO1 questions.

Finally, I wished to consider what features of retrieval practice might be most suitable for a study of this nature, considering the timescale for the work. Through this literature review, researchers have used spaced learning and interleaved practices, leading me to propose that creating a set of low-stakes quizzes, and revisiting topics studied over the A' level course, would be an effective strategy to use in my classroom.

Whilst this intervention has therefore been carefully planned, I appreciate that there are many other factors which may affect the success of the intervention, such as, the views and actions of the teachers delivering the intervention, the timing of it and the other processes and practices within the wider context of the college. In drawing any conclusions from the data collected in this project, such factors will be taken into consideration.

The research questions developed as a result of the needs of the project were formulated as follows:

1. How do teachers and students within the college currently make use of retrieval practice?
2. What is the impact of an intervention on retrieval practice on student performance outcomes in AO1 questions?

Methodology

Practitioner research

The project describes a small scale intervention that has been trialled with two classes of Year 2 A' level chemistry students. Analysis of the impact of the intervention are measured through comparison of student performance in AO1, AO2 and AO3 questions following the intervention. Academic texts written to support planning for research, for example Johnston (2012), Koshy (2010) and Thomas (2009), describe the development of the paradigms that have been used to develop this project. Koshy (2010) described that the purpose of action research is to produce 'practical knowledge that is useful to people working in the everyday conduct of their live and work' as well as 'towards practical outcomes with new forms of understanding'. The outcomes of this study will be fed into work within the department over the following year, building into an action research cycle, however, at the stage of reporting findings this has been conducted. Therefore, this project fits the description of practitioner research, where I, the researcher, am employed within a professional capacity, as a teacher, at the college within which the research is being carried out, (Cochran Smith & Lytle, 2009; Fox, Martin, & Green, 2007; Menter et al. 2011).

Unlike traditional academic research, this project is embedded within my professional practice and that of colleagues participating and collaborating in the research (Cochran Smith & Lytle, 2009; Fox et al., 2007; Menter et al., 2011). I believe that this approach is beneficial for a number of reasons and in my dual role of researcher and teacher, I will be able to interpret data with an understanding of the participants' contributions.

I appreciate that there are limitations to practitioner research. Unlike other scientific research, the data and conclusions generated are specific to my college and the teachers and students involved and cannot be assumed to be transferable to other contexts (Cochran Smith & Lytle, 2009).

However, as the rationale for this investigation is to develop the professional practice within my college, this limitation should not invalidate the findings reported.

In addition, the collaborative element of the research enables me to compare more than one data source which can help assess the reliability of any conclusions that I draw, particularly where my class size is very small, (Cochran Smith & Lytle, 2009).

As I, as the researcher, am working within the college and with the classes used in the intervention, it is not wholly possible for my views to be entirely objective and I recognise that there is potential for bias in interpretation of data collected (Fox et al., 2007 and Menter et al., 2011). Therefore, the methods implemented will be justified and data presented, in order to allow readers to draw their own conclusions as to the degree of subjectivity within the work.

The choices of research methodology based on research studies.

The choice of methodology for this project was influenced by studies on retrieval practice and its impact on student performance. Influential studies, for example Winn (2004), emphasize the importance of monitoring progression throughout a research project, which requires the researcher to be reflective and reflexive. This approach allows the research to be adjusted throughout the project (Johnston, 2012). The intervention was planned to ensure that there were opportunities built into the study to enable strategies to be adjusted as appropriate as the project progressed. For example, the weekly retrieval practice questions were determined in collaboration with a teaching colleague. Where gaps were identified in student understanding during our teaching these could be fed into the weekly quizzes developed.

This project attempts to understand the students and teachers perspectives within a single institution. Data from past years' summer A' level examinations was used to compare outcomes to give some reference points on which to draw additional comparisons in the analysis and evaluation of the intervention. This work therefore will produce context related outcomes sought as a result of the innovation in teaching and learning.

Research instruments

The research instruments used in this study have been summarised in table 2 and include a mixture of qualitative and quantitative methods that reflect the inductive process of this work. Semi-structured teacher interviews gave an insight into their perspectives and the current practices in the department. Student questionnaires collected both qualitative and quantitative data and the rationale behind the methodology is discussed in more detail in the following section.

Table 2 Methodology choices

Research question	Methodology used	Rationale
<p>How to teachers and student currently make use of retrieval practice</p>	<p>Semi- structured Interviews with teachers</p>	<p>Semi- structured interviews were used to collect data about the teacher’s opinions on learning and the use of retrieval practice across the department. As this research question is exploratory and interpretivist in nature, as opposed to seeking a quantitative positivist answer, the ability to ask more probing questions via interviews was considered more advantageous than the larger data sample available by using questionnaires.</p>
<p>What is the impact of an intervention using retrieval practice on student performance outcomes in AO1 questions.</p>	<p>Data analysis of control group and intervention group</p> <p>Student questionnaires* using open and closed questions.</p>	<p>Quantitative analysis of student performance in mock examinations was compared to data from a control group. An unpaired t-test was conducted to evaluate the significance of the data.</p> <p>A questionnaire was used to gather data about the students’ opinions and attitudes after the intervention. This was used to evaluate the impact of the intervention and outline possible improvements</p>

***Small group interviews had been initially planned to collect data, but the methodology had to be altered due to school closures.**

A mixed methods approach was used to answer my research questions, (Menter et al, 2011), in an attempt to collect data to triangulate any conclusions that might be drawn and a multimethod approach in collecting data and evidence was employed based on the work of Schatzman and Strauss (in Wellington, 2000: p23).

To evaluate the impact of the intervention I chose to gather both qualitative and quantitative data, as supported by Cohen et al (2007: p 321) who stated that:

‘Where measurement is sought then a quantitative approach is required; where rich and personal data are sought, then a word-based qualitative approach might be more suitable’.

The use of both qualitative and quantitative data could be used to answer my research questions in different ways. Qualitative methods were chosen to allow for inductive processes for analysis, allowing for variation in setting and participants to be included in the analysis (Marsh, 2005). with the aim of measuring the impact of retrieval practice on students’ attainment post intervention. Quantitative methods enable: ‘patterns to be observed and comparisons to be made’ (Cohen et al. 2007: p 321) and I therefore selected to use quantitative data analysis as a tool to compare student performance.

Through this project I also aimed to gain a more detailed understanding of teacher and students’ thoughts regarding learning and attainment and students experience of the intervention. With these aims in mind, I used a combination of questionnaires and semi-structured interviews to gather data.

Obtaining teacher views on retrieval practice through interviews:

To evaluate how teachers and students within the college make use of retrieval practice at the start of the project I chose to use a series of 1-1 interviews with colleagues across the science department. Interviewing was chosen as the method to collect data as a method to enable me to be able to discuss the participants views and opinions with them during the interview itself and enable me to probe deeper into participants responses, in a manner that would not be possible through the use of a written response questionnaire, (Tuckman & Harper, 2012).

The use of interviews to collect teacher views on the use of retrieval practice would enable me to explore answers given by participants in more detail than a questionnaire. A semi-structured interview design was selected (Denscombe, 2007), which allowed me to steer the discussion via pre-prepared questions (Cohen et al., 2007; Wilkinson & Birmingham, 2003), in comparison to a more structured interview which may have limited replies. This design also provided flexibility to explore interviewee answers in more depth and to probe ideas to encourage interviewees to elaborate on their answers (Marvasti & Freie, 2017).

Due to the time constraints of the project, it was not possible to pilot the interview questions beforehand. To compensate for this, the 1st interview was conducted and analyzed in detail before subsequent interviews were carried out. This enabled me to consider the wording used in the questions to enable me to collect comparative responses from each participant enabling comparisons to be made. Using Patton's (1980, p 206) analysis of the strengths and weaknesses of different types of interviews I adopted an interview guide approach which enabled me to increase the depth of the data collected. The interview questions were determined in advance and reviewed and modified with an expert before being used. The validity and clarity of the items were discussed to increase the comparability of participant responses and in an attempt to reduce interviewer effect and bias. Four key questions were designed to prompt discussion and the guidelines on prompting by Denscombe (2007) were used.

The interviews were carried out on a one-to-one with volunteering participating teachers and the purpose of the interview was explained in advance. The teacher interviews took place in the teacher's own classroom space. This was selected as a comfortable space (Koshy, 2010), in an attempt to reduce the power balance between myself and my colleagues. Volunteers were used in an attempt to reduce the power balance between myself and my colleagues due to my SLT role within the College. (Farrimond, 2017). To encourage relevant discussion, pre-prepared interview questions were developed based on key factors identified in the literature review. Open questions were used to give the teachers opportunities to discuss to questions and generate discussion.

I selected to record the interviews and transcribe them. This enabled me to conduct the interviews as a conversation between myself and the interviewee, to explore their responses in more depth and encourage the flow of conversation. Although the time needed to transcribe recorded interviews is a disadvantage of this method of data collection, (Cohen et al., 2007, p271), the transcription process enabled me to reflect on the answers given and revisit them to compare and contrast participants views.

In the analysis of the responses to the interview questions the teachers have been given a code so that they cannot be easily identified from their answers in an attempt to anonymise their responses.

The interviews followed the format described by Marsh (2005 p108) with the participant taking the discussion. I was aware during the interview that it was very difficult not to lead the participant in their answers and that it was important not to 'fill' the gaps. I attempted to give each respondent enough thinking time to respond to my questions.

Due to the size of my college and department, my sample size excludes causal analysis (Borg & Gall, 1979, p 194). However, each participants has recently taught in a range of other academic establishments and could make comparisons between more than one teaching environment.

In addition, their responses are relevant to the situation that we work in and the sample size, whilst small, did cover each science discipline, biology, chemistry and physics across the college, and was considered suitable for the purpose of this study.

The use of quantitative data analysis to evaluate the impact of the intervention on student performance

A testable hypothesis was developed to answer this research question due to the positivist nature of the question. The question seeks to identify relationships between variables (use of retrieval practice and student performance), using quantitative data analysis (Thomas, 2009) with a successful innovation hypothesised to increase student performance in AO1 questions. The quantitative methods used in this work were based on those described by Wellington (2000, p 17) in addition to the guidance described by Braun, Clarke & Gray (2017) of selecting techniques in terms of the research questions being asked. In addition, when selecting appropriate methodology for this study, consideration was given to specific ethical issues presented. As all data collection methods present their own methodological advantages and disadvantages, these were considered to determine the most appropriate methodology for this project.

The impact of retrieval practice evaluated by comparing the mean score in AO1 questions of the control and the intervention group and an unpaired t-test, appropriate for small sample sizes, was used to determine the significance of the findings, (Corder & Foreman, 2014, p 244). The use of retrieval practice may be considered to be successful if there is a statistically significant difference between the performance of the two student groups.

Obtaining student opinions about the intervention

To explore causality, in the original planning stages of the project the intention has been to carry out small group interviews with the intervention group of students to determine their views on it. Due to the Covid19 pandemic national lockdown alternative forms of data collection had to be considered as face to face interviews were not possible. The use of Internet based methods of communication, such as Skype, FaceTime and Zoom), enable interviews using voice and video via a real-time connection. Such methods of communication are becoming increasingly important and influencing researchers' options, enabling researchers to contact participants worldwide in a time efficient and financially affordable manner, (Braun et al 2017). Such technologies, therefore, have the potential to increase the variety of samples, although recent research by the Sutton Trust (2020), highlighted that the poorest in the community have less access to such technologies and may be precluded from studies.

Internet based methods of communication would have enabled me to conduct face to face interviews with students, and therefore the small group interviews initially planned could have been conducted. However, whilst my students are able to use such technology confidently in some cases their internet connection is not reliable. Braun (2017) concluded that video-calling is a close substitute to in-person interviewing and would allow data to be collected over large geographical areas, which is of particular benefit to me as the international status of the college means that students were now distributed across the globe. However, in my experience of teaching online live lessons using such technology via Microsoft Teams and Zoom platforms I have found that areas of rapport and the lack of verbal cues that exist in face to face teaching were problematic. Therefore, although conducting qualitative interviews using such technologies initially looked promising, I concluded that the lack of face to face interaction and that I would miss nuances in students responses due to the lack of ability to read their body language. I was also concerned that such methods would allow more confident students to dominate the interviews and that I would not hear the voice of all students.

There are also additional ethical issues associated with the use of such technologies, that had not been considered when applying for CUREC at the start of the project. In terms of safeguarding, the college advises that all such meetings conducted via such technologies are recorded. This would lead to issues in terms of ensuring anonymity and also data storage. Students may also be reluctant to openly answer questions when aware that the interview was being recorded, or give the answers that they thought that I wanted to hear. Therefore, a more traditional survey style questionnaire was designed to elicit student perceptions about the intervention and due to the closure of the college it was necessary to conduct an online survey to collect data. The benefits of using a questionnaire to collect data, (Wilson and McLean's (1994), outweighed the closed nature questions and their responses at this stage of the data collection.

Student questionnaires

To review the impact of intervention on retrieval practice on student performance and to evaluate the intervention from their perspective, students in the intervention group were asked to complete a questionnaire. Response rate to questionnaires has been reported to be very low (Bell, 1999), but the purpose of the questionnaire was explained to the classes who had participated in the intervention in an online lesson. The students were made aware that the data collected would form part of this work and by involving them in the project I hoped that the response rate would enable conclusions to be drawn. The questionnaire was designed with a limited number of questions to minimise the time needed to complete the survey in an attempt to encourage more responses from students, to enable conclusions to be drawn from their responses.

In light of having to make changes to the methodology at short notice and also the fact that my students are EAL and cannot always express themselves coherently in open-ended questions, I concluded that the use of closed-questions in the questionnaires was appropriate at this stage of the project.

Closed questions do not allow respondents to add any extra detail or qualifying remarks to their answers (Oppenheim, 1992: p 115) and may also encourage respondents to give polarised responses, (Koshy, 2010). However, the use of closed questions in my questionnaires enabled me to generate quantitative data which I used to identify trends and patterns in the students' responses. I also provided the option for students to provide an explanation of their responses, if they wished to. The intention of this was to elicit more detailed feedback from students that could not be measured on a point scale.

Ethical considerations

The ethical considerations for each of the methods outlined were discussed with the project supervisor before any of the action research cycles and any interactions with the college were carried out. Then British Educational Research Association set out guidelines in 2011 which were read and understood (BERA, 2011). In addition, the University itself has a methodical approach to approving project work in schools, which has to be seen and passed by the Central University Research Ethics Committee. Thus, alongside the usual professional ethical standards upheld by teachers, and in compliance with local government laws, such as the GDPR, the ethical considerations identified by BERA (2011) were thoroughly discussed and adhered to throughout the project.

Procedures put in place required the anonymity and confidentiality of information collected and no mention is made of the name of the college or any participant throughout the project or the reporting thereof. All participants were treated with respect and their consent was sought before any activities were carried out at all times. They also maintained the right to withdraw their consent for participating in the study at any time and would have their data removed, returned (or destroyed - as appropriate) from the study data. The consent of the college was also obtained before the project was carried out as part of the CUREC process. The Principal and Board of Directors were exceptional in their support of the work and were keen to participate and learn from it. This meant that the leadership of the school were also keen to oversee and participate in the activities going

on in the project when they could. Indeed, at times, some of the senior members of the school leadership team were actively involved in the research activities themselves and reported to have gained much from the experience and were satisfied that the activities were carried out with due regard for the ethical obligations of the collaborators involved.

The interview recordings collected were kept in line with the school ICT acceptable use policy, which required them to be stored on the secure internal servers and iPads used were password protected. All devices were kept securely on site at all times.

Examples of student's work was scanned unnamed and stored in on the same secure internal servers. Recording of staff interviews were only analysed by myself and discussed only with my supervisor. Recording have not been heard by anyone else, nor will be used for research without further explicit permission.

Finally, all participants, including students and teachers have been anonymised in all research reports. The data collected has been be kept strictly confidential, available only to my supervisor and the researcher, and not used other than specified without further consent. All recordings will be deleted at the end of the research cycle.

Throughout the project I was aware of a key area of concern relating to the power relationship between myself and the participants in the study. This is in terms of the teachers and my role both as a researcher and on SLT and in terms of the students in my role as both the teacher and researcher. Such power relationships may be considered to increase an already asymmetrical relationship within the interview process (Farrimond, 2017). Due to the small number of students in the college, it was necessary to carry out the research with students that I teach. However, through collaboration with a fellow chemistry teacher the intervention was also conducted with those who I do teach in an attempt to triangulate conclusions drawn.

Following permission from the Principal to carry out this research, I obtained consent from all participants, both students and teachers. Consent was informed (Farrimond, 2017) and was obtained in the form of process consent. Consent was obtained at each stage of the research process, thus it could be withdrawn at any time. Participants actively consented to each stage of the process (Farrimond, 2017).

The ethical considerations, in terms of collecting data during a global pandemic, were not considered in the initial CUREC application for this work. However, whilst it was necessary to adjust the initial aims to conduct face-to-face interviews with students, and move these to online questionnaires, at all times during the data collection I considered whether asking students and teachers to participate in research at this time would put them under any additional unnecessary stress. The health and wellbeing of participants was considered at all stages of the data collection, in addition to safeguarding considerations when moving from face-to-face to online data collection. The selected methodology fits traditional methods of data collection as described in the original CUREC application and therefore it was not considered necessary to apply for further ethical consideration during the project. Prior to sharing the questionnaire, the purpose of the investigation was outlined to the students. I explained how the results would be used, and emphasised that the questionnaire was both optional and their responses completely anonymous.

The participants in the study

The intervention group was made up of two classes of A'level students, taught by two teachers in the college. N = 16, with a 50% male : 50% female gender balance; mean age at end of the A'level 19.12 years. Average % score on baseline test at start of the course 66.1%.

The control group gender balance was 60% male and 40%; mean age at end of the A'level 19.36 years. Average % score on baseline test at start of the course 65.4%.

The teaching participants in this study were 3 teaching colleagues, 1 female and 2 male teachers. They have, on average 10 years teaching experience across a variety of different school types and participation was voluntary.

The format for obtaining teacher-researcher collaborators

An invitation was extended to colleagues in the science department during a departmental meeting to collaborate in this project. One chemists, one member of the biology department and a physics teacher volunteered for 1-1 interviews. All members of the chemistry department undertook to trial the intervention strategy with their classes and to collate and share data. The teacher of chemistry was asked to also be a 'critical friend', as suggested by Koshy (2010) and assisted with the technical development of the tasks. She also supported with defining the subjective and syntactic development for which the other members would not be qualified to assist with in this project. A regular meeting time and place was agreed to review the project, after consent was sought from all participants and the ethical considerations were clearly described to the participants.

The intervention

From my literature review of the use of retrieval practice strategies, I evaluated the methods suggested to develop a classroom intervention that would be most appropriate for my students and their needs. The intervention was based on developing AO1 questions where my students were required to actively retrieve and write the answers before they could then self-check them. Low-stakes quizzes were designed to be conducted weekly, at the start of a lesson. The rationale behind the quizzing was explained to students at the start of the intervention. The students were taught to answer the questions independently and in silence, to enable them to attempt to actively retrieve in the information themselves without distractions from others.

The series of quizzes, in the form of worksheets, were developed using AO1 questions as identified by the exam board and using platforms, such as Edexcel's exam wizard, as a search engine. I decided to produce my own quizzes for the intervention so that I could tailor them to the needs of my own class and cohort. The quizzes consisted of 9 questions with typical marks ranging between 14-18 marks. This number of marks allowed students to take the quiz at the beginning of a lesson, mark it and then use the remainder of the lesson to continue the teaching programme. Students were asked to correct their answers in their own time and to record their own thoughts about what they needed to do to improve during the feedback process. An example of a typical quiz is shown in figure 5.

11

Describe the nature and strength of the bonding in magnesium oxide. MgO Ionic bonding Strong electrostatic forces between oppositely charged ions. (attracted) 3/3	Define the term mass number Number of protons and neutrons 1/1	What are isotopes? elements with the same number of protons, but different number of neutrons 0/1
What is a primary alcohol? Alcohol, in which carbon, attached to OH group, has only one carbon attached to it 1/1	Correct this equation for the 1 st ionisation energy of sodium $Na - e^{-} \rightarrow Na^{+}$ 2/2	What name is given to this equation: $PV = nRT$ Ideal gas equation 1/1
Calculate the amount, in moles, of magnesium in a 0.0732 g sample. Mg $n = \frac{m}{M} = \frac{0.0732}{24.3} = 3.01 \times 10^{-3} \text{ mol}$ 1/1	Write the ionic equation for the reaction between sodium hydroxide and hydrochloric acid. $NaOH_{(aq)} + HCl_{(aq)} \rightarrow NaCl_{(aq)} + H_2O_{(l)}$ $OH^{-}_{(aq)} + H^{+}_{(aq)} \rightarrow H_2O$ 1/1	Define standard enthalpy of formation of MgO The energy change making 1 mole of MgO from its elements: magnesium and oxygen under standard conditions 2/3

Score 12/18

Figure 5 An example of a typical quiz

Open response questions were selected for the intervention tasks rather than multiple-choice responses, based on the findings of Butler and Roediger, (2007), Kang et al., (2007) and McDaniel et al., (2007).

Although multiple choice questions would allow for easy peer-assessment and make them appear less threatening to the students, I wanted to remove the possibility for students to 'guess' the answer and to have to more actively retrieve information rather than 'recognise' the correct answer from those provided. I also hoped that in having to formulate their own answers, that the quizzes would additionally support my EAL students by enabling them to practise presenting their thoughts on paper in a low-stakes environment.

An additional consideration of the paper-based tasks and the design of the intervention was to support the learning of EAL students and help them to develop their reading skills, which vary across the cohort. Learning walks across the college focussing on the teaching of EAL students have previously highlighted that they spend a considerable amount of lesson time listening. Therefore this intervention was designed to provide an opportunity for students to work independently and read and process information as they would be required to in an examination.

In introductions to the intervention, students were told that the marks would not be collected or used in anyway by the teacher, and that they should be prepared to think hard to try to recall the information from memory, without referring to class notes. If students were unable to answer a question, they were encouraged record this, and for this to become their own personal homework after the lesson.

Students were provided with an individual copy of the questions, stuck into an exercise book, as they arrived at the lesson and wrote down answers at their own pace within a time limit, answering and then checking their work independently. In lessons, students choose between using paper and folders or writing directly to an electronic device to record their notes. By choosing to have an exercise book for the tasks, I wanted to make this type of task stand out as different to normal working practice. The weekly quiz questions sheets were stuck into a dedicated book for retrieval practice. The purpose of this was to provide a revision tool for students at the end of their course and to provide a space for students to correct their own work.

Karpicke (2009) has reported that ideally a fact should be successfully retrieved three times before a student moves on from it, so questions used in quizzes were repeated over the course of the intervention so that the majority had been seen at least three times by the students over different weeks.

Feedback and marking of quizzes

The answers were shared on the classroom whiteboard, and when they were ready students self-checked their work. The aim of self-marking their own questions was to encourage students to evaluate the depth of their own learning and to set their own learning goals. In the feedback session, common incorrect answers were discussed. The aim of this process was to keep the outcome of the assessment with the student so that they learnt what they did and did not know. This intervention strategy was designed to enable me to be free to circulate and spot common errors as they emerge, which were brought into discussion during the whole class feedback.

In feedback given at the end of the retrieval quizzes, students were encouraged to see mistakes as opportunities to learn, to emphasise the low-stakes nature of the quizzes, asking students to view mistakes as a natural part of learning. In my literature review, studies had shown that allowing students a role in setting timelines for completing tasks and monitoring their own progress helps them to acquire mastery and therefore improve performance (Poortvliet et al. 2007), and this strategy was employed in this intervention. Students were then asked to identify how they could improve their answers and to work on them in their own time before the next quiz. Examples of how students did this are shown below, figure 6.

6

If hot
 $\text{NaCl} + \text{NaClO}_2 + \text{H}_2\text{O}$

<p>Which of these solutions does not contain the same total number of ions as the others?</p> <p>10.00 cm³ of 0.100 mol dm⁻³ NaCl(aq) 20.00 cm³ of 0.050 mol dm⁻³ NaCl(aq) 20.00 cm³ of 0.050 mol dm⁻³ MgCl₂(aq) 13.33 cm³ of 0.050 mol dm⁻³ MgCl₂(aq)</p>	<p>Sodium nitrate decomposes on heating.</p> <p>$2\text{NaNO}_3(\text{s}) \rightarrow 2\text{NaNO}_2(\text{s}) + \text{O}_2(\text{g})$</p> <p>What is the maximum volume of oxygen obtained by heating 0.50 mol of sodium nitrate?</p> <p>[Molar volume of a gas = 24 dm³ mol⁻¹ at room temperature and pressure]</p> <p>$0.25 \times 24 = 6 \text{ (dm}^3\text{)}$</p>	<p>What are the products when chlorine is passed through cold, dilute aqueous sodium hydroxide solution?</p> <p>$\text{NaOH} + \text{Cl}_2 \rightarrow \text{NaCl}$ $+ \text{NaClO} + \text{H}_2\text{O}$</p> <p>Basic solution HClO can't be formed</p>
<p>What is the expression for ΔS_{Total}?</p> <p>$\Delta S_{\text{surroundings}} + \frac{\Delta H}{T}$ $\Delta S_{\text{surroundings}} - \frac{\Delta H}{T}$ $\Delta S_{\text{system}} + \frac{\Delta H}{T}$ $\Delta S_{\text{system}} - \frac{\Delta H}{T}$</p>	<p>Which of the following ions has the biggest radius?</p> <p>S^{2-}, Cl^-, K^+, Ca^{2+}</p>	<p>Draw a diagram to show the bonding in the single product of the reaction between BF₃ and NF₃.</p>
<p>Sodium chloride is ionically bonded. What is meant by the term ionic bond?</p> <p>The bond between two ions: metal-non-metal bond. Elements should have big difference in electronegativity</p>	<p>Give the formula of an ion from Period 3 that is isoelectronic with the chloride ion.</p> <p>S^{2-} P^{3-}</p>	<p>Complete combustion of 50 cm³ of a hydrocarbon vapour gave 350 cm³ of carbon dioxide, both gas volumes being measured at the same temperature and pressure.</p> <p>Which could be the formula of the hydrocarbon?</p> <p>C_2H_6, C_2H_4, C_2H_2, C_2H_8</p>

Electrostatic attraction between two oppositely charged ions

Score 8/12

Figure 6 An example of a student work with self-corrected their answers

Gathering feedback on how well pupils have learned a topic is important in enabling teachers to address any misunderstanding and provide the right level of challenge in future lessons. To ensure that the retrieval quizzes used in the intervention were low stakes, the marks were not collected from the students. They were asked to record their performance and identify areas for development. An example of how they recorded this is shown in figure 7.

Weekly Review Progress Tracker



Week	Score	Feedback / Improvements
1	9/9	
2	9/9	
3	8/9	Li_2SO_4
4	11/16	① HNO_3 : nitric acid ③ Fl and aq is not the same! ② H_2SO_4 : sulfuric acid ④ Crucible: \rightarrow solution.
5	15/16	volumetric flask 
6	10/12	$[\text{Ar}]3d^5 = \text{Mn}^{2+}$
7	10/14	S in Group 6. 
8	8/12	Relative atomic mass: Average mass of an atom compared to $\frac{1}{12}$ of the mass of a C-12 atom.
9	8 10/14	
10	14/18	A hazard is a measure of harm. Br_2 , Hg (l)
11	9/15	$\text{X}_{19} - e^- \rightarrow \text{X}^+_{19}$
12	12/14	As last & first (before 3d) Hess law definition.
13	9/10	Read the Qs <u>Carefully</u> !
14	6/10	Dec - 10. Oct - 8 CIP RULES • Tetrahedral = 109°

Figure 7 – A student record sheet

The vocabulary used was to call the exercises ‘quizzes’ rather than tests. It was explained to the students that that they were designed to help students self-assess topics and areas of the course that they were confident with, and areas that required some development. The students were encouraged to record their own results for tracking and to self-correct their own answers.

The routine ensured a calm start to the lesson. Students were observed to immediately fall into the safety and calmness of the routine. The initial set of quizzes included questions that were structured to set the learning episode up to succeed for as many as possible, with questions based on the last lesson taught.

In general I did not disturb any student during the course of the quiz. Students were left alone with their thoughts and were able to self-check their answers when appropriate for them. I made a judgement for when the retrieval phase ended, normally between 5 and 10 minutes of the start, when all students appeared to have reached the last question.

Feedback was then provided, live and under the visualiser. Students were encouraged to 'mark like a teacher': this is a phrase I used to remind them that incorrect answers are learning opportunities: to focus their attention on the feedback. They annotated corrections for themselves explaining either a correct method or answer. In addition, as part of the lessons planning I considered which questions I intended to explain, which I might ask students about. The feedback was designed to give students an indication of what they had done well and how to improve. By sharing model answers to questions, students would see what is expected of them in an answers, and where marks are gained and lost. They were also exposed to assessment language, for example the difference between 'state' and 'describe'. The feedback method also aimed to encourage metacognition, by encouraging students to think about their own learning and thinking and by aiming to develop skills of self-critical reflection. The aim post quiz was for students to take charge and responsibility for their own learning by setting them the goal of reviewing work independently that was not yet embedded.

The intervention was carried out with two classes of 10 students of mixed ability. There were two classes in total, taught over a 12 week period of five, 60 minute lessons a week. In the normal timetable students have 6 timetabled lessons a day, starting at 08:30 in the morning and finishing at 16:30. The intervention took place during their usual timetabled chemistry lessons.

Findings and discussion

In the initial planning of this project it was proposed that it would be conducted with Year 2 students completing external A' level examinations in the summer term. This would provide data from questions which would enable a comparison of student performance in AO1, AO2 and AO3 style questions both pre and post intervention as a measure of the effectiveness of the intervention strategy. The Covid19 global pandemic and national lockdown of schools and colleges across the UK from March 2020 with the cancellation of all external examinations, meant that the planned data collection was not viable. All lessons within the College were also taught remotely from March 2020.

I therefore had to suspend data collection and re-design part of this project to take into account social-distancing measures. Therefore the data forming the analysis consisted of that of the control group whose performance in the A' level chemistry examinations in 2019 led to this project, and mock examination data collected for the Year 2 students in the 2020, who were given the same exam paper sat by the original cohort.

There were some advantages to this enforced action, in that the paper used in the mock examination was identical to that sat by the control group and therefore direct question by question analysis was possible. This data analysis would not have been possible if the 2020 examination series had gone ahead. However, students in the intervention group sat the mock paper several weeks earlier than the proposed A' level examinations would have taken place and prior to participating in a structured revision program that had been completed by the control group. In addition, the mock examinations were not sat under the same examination conditions as the paper sat by the control group and some students may not have taken the mock as seriously as a 'real' examination. A number of studies have reported a clear difference in motivation when the test-stakes are changed. Wolf et al. (1995) have reported an effect size for motivation of 1.45 between the two test conditions, where one examination was low stakes and another contributed towards a student's final grade.

In a similar study by Sundre (1999), a smaller, but still significantly positive effect size of 0.79 was found in the motivation reported between two groups. Additionally, a difference in outcome for more cognitively demanding questions, 1.38 compared to 0.62 for multiple choice questions, was also supported in the findings of Sundre and Kitsantas (2004) and Barry et al (2010). Therefore in analysis of data collected and comparison between the control and intervention group it is necessary to consider that increased motivation results in increased performance and draw conclusions with this in mind.

A network analysis (Bliss, Monk and Ogborn, 1983) was used to analyse the qualitative responses for both the teacher interviews and student questionnaires. Reading through the student responses and listening to the teacher interviews several times, enabled me to identify themes that emerged. Ideas were ranked in terms of the frequency with which they were cited, which allowed me to identify the themes were considered to be the most important by the respondents and to see patterns in responses given.

Data analysis

RQ 1 - How do teachers and student currently make use of retrieval practice?

Face-to-face interviews were carried out with teacher colleagues within the college, as part of the collaboration aspect of this work, prior to the college lockdown. Analysis of these interviews are described here.

To answer the 1st research question, interviews with teaching colleagues were conducted using a semi-structured interview approach as discussed in the methodology. A number of themes emerged from analysis of the interviews and are discussed here. Quotes used from the interviews have been chosen as they answer the research question I am exploring.

During the interviews, the teachers interchanged their use of the word knowledge and understanding and used both words when discussing the knowledge-based content of their courses. There were also instances where throughout the interviews, the teachers contradicted themselves. Sometimes they acknowledged the contradictions and rephrased answers to earlier questions, but at other times an answer to a follow up question contradicted an earlier response. The interviews have been analysed with these observations in mind.

The themes that emerged from the teacher interviews were

- That subject teachers believed that there is a certain proportion of material within their discipline, that needs to be learnt in order to be able to understand later concepts.
- There was an assumption across the department that students would learn material presented to them routinely.
- That learning subject knowledge was relatively trivial.
- That, at the start of the project, teachers knowledge about retrieval was limited.

These themes are described further below.

Knowledge base vs skills base

At the start of the project I considered that the more subject knowledge needed, the more applicable retrieval practice would be to that subject and that more likely that the subject teacher would see the benefit of this project. Therefore, the initial question asked in the teacher interviews was to determine what proportion of their subject they felt was knowledge based and what proportion was skill based. The aim of this question was not to get an absolute value for each subject, but to encourage the interviewee to begin to consider the relative amount of learning that was required within their subject, so that subsequent questions were answered with this in mind. The question was also aimed to begin to get the teacher to consider some of the factors that might influence student performance in examinations.

As the meaning of skills and knowledge meant different things to different teachers, I explained to teachers at the start of the interviews, that for the purpose of the interview I was defining knowledge as the subject content whereas skills included practical skills, thinking skills and other subject specific skills as described by Petty (2006). The exact definition of what was meant by 'skills' was not considered to be important at this stage in the interview. The term 'skills' is used routinely within the department, relating to practical skills students gain through their courses. If I were interviewing teachers from outside the college, time would be given to sharing our understanding of terminology.

By asking participants to consider what proportion of their subject was knowledge based compared to skill based, I aimed for the teachers to consider what students need to 'know' in order to be successful in that subject, in order to then apply it, in practical work or examination questions. The interviews highlighted the consensus between the teachers that when students were presented with an exam paper, the majority of questions required them to be able to apply learnt knowledge, although the amount of subject knowledge and subject content might vary in proportion between the three sciences. Examples of where students need to know key ideas before they can answer questions were given in all of the interviews.

In chemistry and biology, whilst there were differences in opinion regarding the relative amount of knowledge thought to make up the courses, the teachers agreed that there was a significant amount of subject content that the students needed to be able to draw upon. The biology teacher commented that "there is a huge amount of knowledge that you have to put across in biology" and the chemistry teacher stated that

"yes, yes they do [need to know key facts], there are fundamentals in chemistry they need to understand and know before they can apply anything".

The physics teacher initially said that he considered physics “to be more skill based than a lot of other subjects which are taken at secondary education”. When this answer was explored through further questioning he said that “if you haven’t got secure subject knowledge then you can’t apply it”, seemingly contradicting the amount of subject knowledge compared to skills that he had initially indicated. In an exploration what he meant by ‘secure subject knowledge, he developed his answer saying that

“there are certain areas of the courses where rote learning is needed”. He said that the “exam tests the ability of student to apply it [knowledge] rather than just their memory of an equation”

adding later that

“students need to know key facts before they can answer a question - there is a certain level of knowledge that is needed, for example a relevant law, before they can apply it”.

The interviews all established that in each of the different subjects that there was a certain amount of knowledge that students needed to learn in order to be successful in answering exam questions.

All of the teachers interviewed agreed that any form of rote learning was not enough to enable students to secure the top grades in their subject. A theme that emerged from all of the interviews was that teachers across the college and science department were not driven by exam performance. A philosophy of instilling a love and passion for the subjects and an understanding of them came through in all of the interviews. From responses given, the college ethos of developing students with higher order thinking skills and learners that have a mastery of the subject was a key aim that the teachers said that they tried to achieve through their teaching. Recognizing that students are tested, and their performance measured against national standards at the end of the courses was however recognized and acknowledged.

How do we, as teachers, know that learning has taken place?

In the interviews the teachers were asked to describe how they know that their students has learnt key ideas and concepts and to explore strategies that were used by teachers to encourage learning. It was common across the responses given that the teachers thought of learning as expecting students to be able to recall ideas without reference to any notes or books. For example, one respondent said that it was “that information that they could remember or recall without reference to learning aids such as textbooks or class notes”.

In exploring how they thought that their students went about learning the material that the teachers considered to be core knowledge and what strategies that they used to promote learning, a theme that emerged from the responses was that throughout the department, learning was something that was assumed to be happening. In the interviews, all of the teachers recognized that they made the assumption that their students would have learnt the materials presented in previous lessons, without the teacher having explicitly instructed them to or given them support in how to learn the material.

When asked how the teachers believed that their students learnt the key ideas that the teachers had identified, the interviewees tended to talk about what they did to promote and encourage learning, but none discussed the student’s role in their learning. None of the interviewees could give examples of what they did to encourage the students to learn material, nor how they knew or assessed if learning had taken place. When prompted, the teachers talked about their use of question and answers to check a student’s understanding but, as in my lessons, there was no established approach to check if the students’ knowledge was secure at the start of the project.

Another key theme that emerged, across the department, was that learning key ideas such as definitions was considered to be a trivial process, despite the evidence from the AO1 scores showing that this learning was not taking place. For example, one teacher commented that students “just need to remember the definitions” implying that learning was a simple process.

In a separate interview, another teacher commented that the rote learning of key facts “is something that the weaker students do”. All of the teachers used the term “simple rote learning” at some point during their interviews, implying that they considered the learning of definitions and key ideas to be trivial.

In follow up questions no evidence could be provided by the teacher to support these statements, which again reinforced my observations that as teachers in the department we were assuming that learning was happening without checking that it was taking place. In exploring how teachers knew that learning was taking place one teacher commented that “I don’t specifically test it, I would rather keep the focus [in lessons] on application”. He later commented that

“the message I give them [the students] is that they shouldn’t open the data booklet [provided in the exam]. It is there as a safety net”

again highlighting the assumptions that students do the ‘learning’ (in this case of equations) independently, without being directed to do so.

Teacher’s Prior knowledge about retrieval practice.

Interview questions were used to elicit what knowledge of retrieval practice teachers in the department at the start of the project. The questions were then developed to gain an understanding of approaches that teachers and students used that could be described as retrieval practice.

When asked what the teachers knew about retrieval practice at the start of the project, prior to any intervention being carried out with students, the responses showed that none of the teachers were familiar with the terminology. One interviewee replied, “not a lot really to be honest” and another “I guess I don’t really know”. However, on following up on their responses, it emerged that the teachers did employ strategies that could be classified as retrieval practice.

Exploration of approaches used in the teaching classroom room showed that techniques such as interleaving and spaced practice were occurring, but these were not recognized as being retrieval practice or formalized in any way. All interviewees discussed their use of past paper questions in lessons, and the idea of repetition of topics and ideas to reinforce concepts taught came across in all interviews. On further questioning it emerged that the use of past paper questions was not consistent across the science department with different practices in different subject areas. In some cases, teachers used them as homework from early on in the course on a weekly basis whereas in other subjects they did not do any past paper questions at all until the end of the taught material in Year 2 of the A' level course.

At the end of the interviews all of the participants commented that they had forgotten that the interview was being recorded and I observed that the interview flowed more like a conversation between colleagues. Therefore, I feel that the responses given reflected the true opinions and were not influenced by either my role as SLT or by them feeling that they needed to give a 'right' answer.

The interviews highlighted that learning was considered to happen without explicit support or guidance and no formal practices were in place to check that this was indeed the case. These observations are in line with those reported by Binks (2018) and the use of testing to support and enhance learning was limited.

The impact of retrieval practice on student performance in AO1 questions.

The intervention group sat A' level papers from the 2019 summer examination series as mock examinations. Questions from the papers had been categorised as AO1, AO2 or AO3 by the exam board. Student performance in the different types of questions was analysed and compared to the performance of the control group, who sat the same paper the previous summer. This allowed for direct question by question analysis to take place. The average performance in AO1, AO2 and AO3 questions of the two groups are shown in table 3.

Table 3 Average score in AO1, AO2 and AO3 of control group and the intervention group

Assessment objective	Mean score Control group	Mean Score Intervention Group	Improvement
AO1	64	83	+ 30%
AO2	70	87	+ 24%
AO3	60	81	+ 35%

The data shows that the intervention groups average score for AO1 questions increased from 64% to 83%. The average score for students in AO2 questions increased from 70% to 87% and in AO3 questions from 60% to 81%. The difference pre and post intervention showed an average of 30% increase across all assessment objectives. The data is represented graphically in figures 8-10.

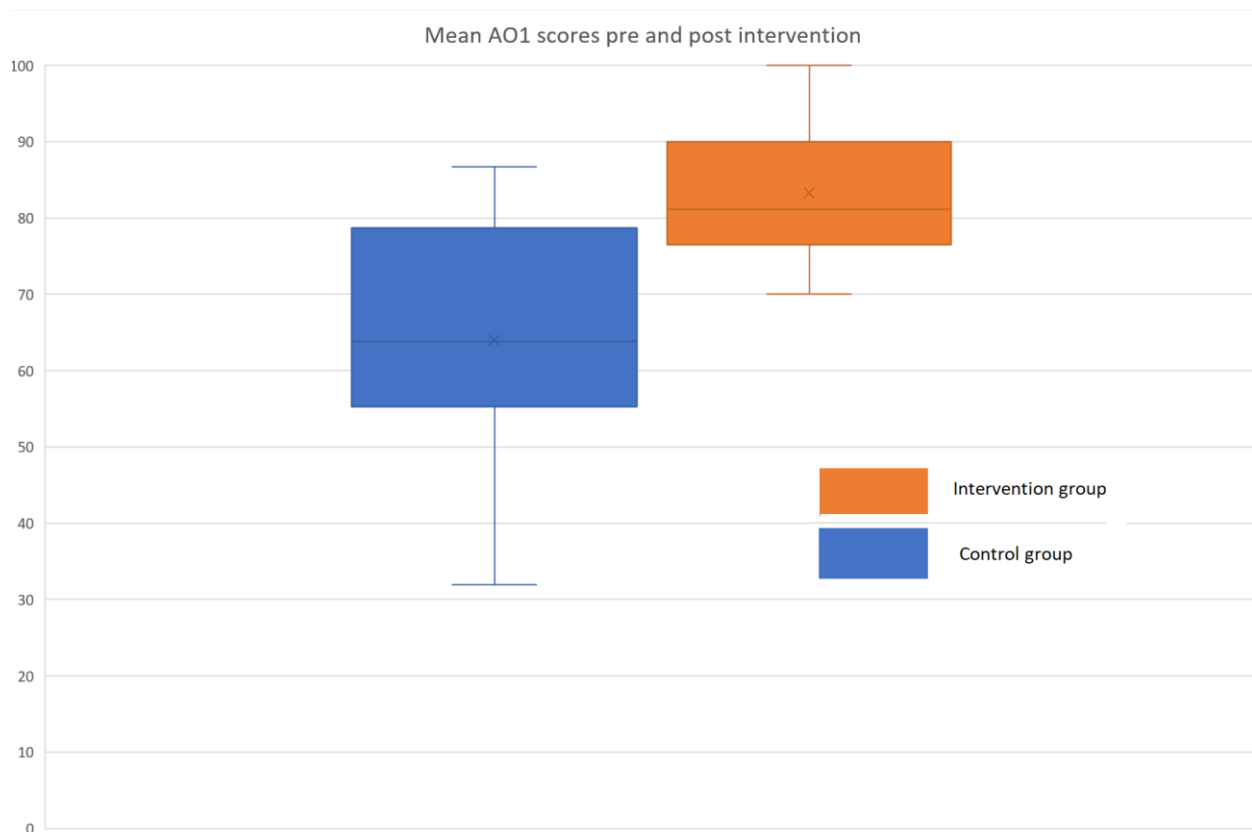


Figure 8 Comparison of student performance in AO1 questions pre (blue) and post (orange) intervention

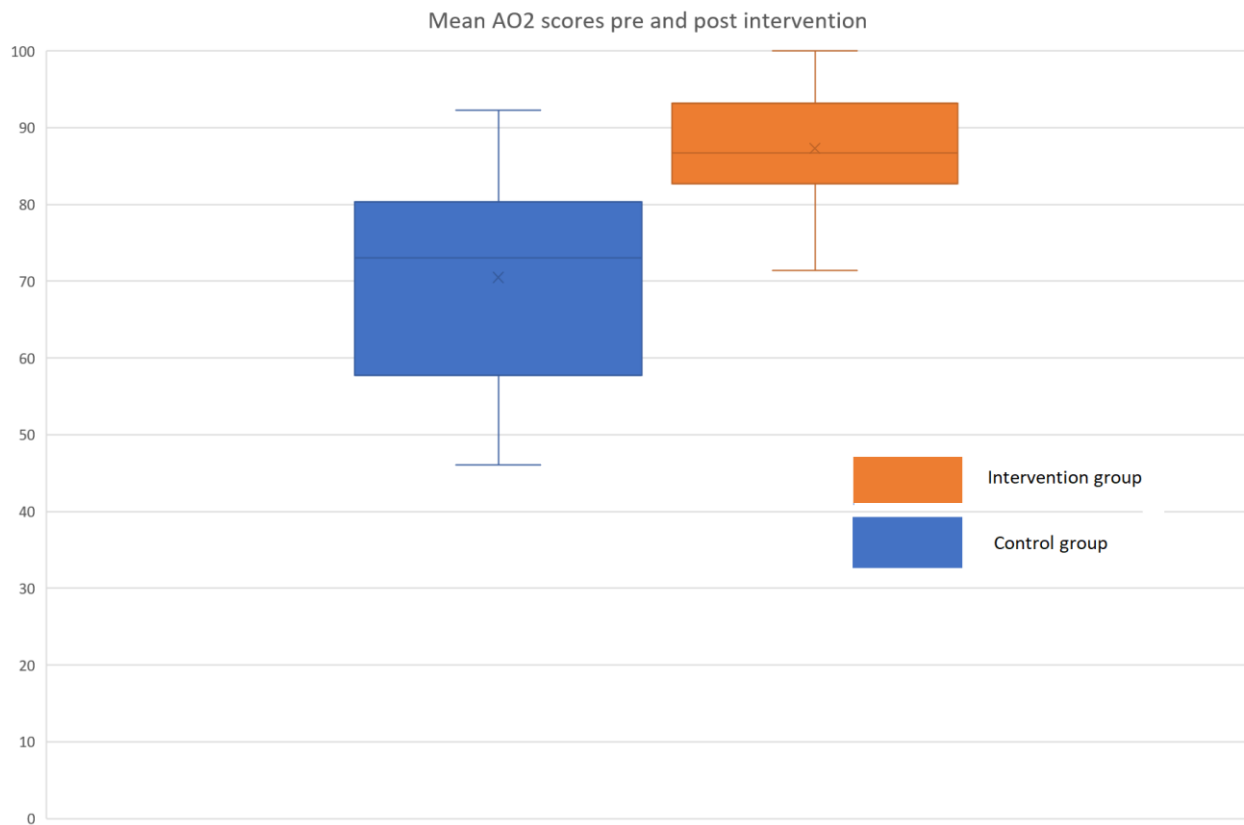


Figure 9 Comparison of student performance in AO2 questions pre (blue) and post (orange) intervention

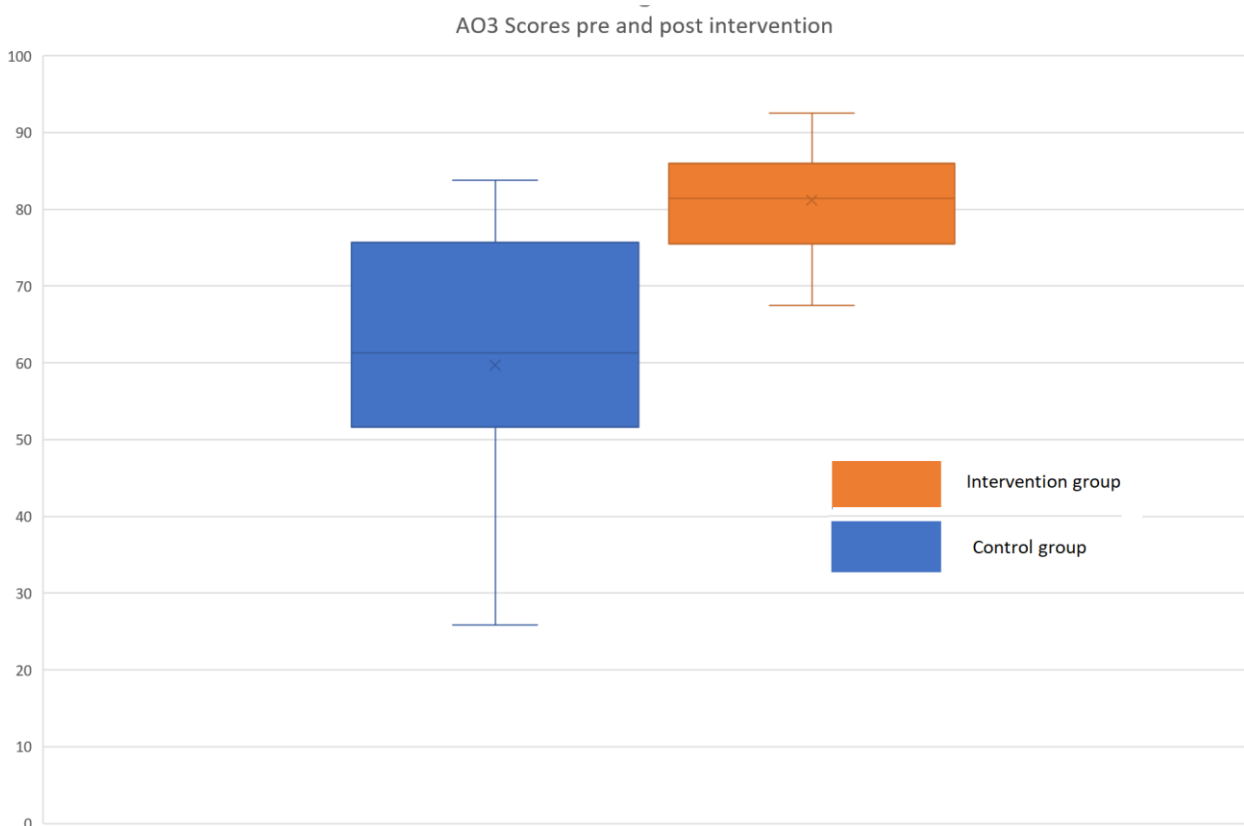


Figure 10 Comparison of student performance in AO3 questions pre (blue) and post (orange) intervention

To evaluate any significance in differences between the performance of the two groups, an unpaired t-test was conducted, comparing the mean and standard deviations between the data sets.

The unpaired t-test analysis showed that the difference in student performance post intervention may be considered to be statistically significant, with a 95% confidence interval, for all of the assessment objectives, (AO1: $t = 4.28$, $df = 31$, AO2: $t = 4.22$ and AO3: $t = 4.52$).

The unpaired t-test analysis showed that the difference in student performance post intervention may be considered to be statistically significant. Thus there is evidence, at the 5% level of significance, to suggest that the intervention group did better than the control group in all three assessment objectives. This analysis suggests that the use of retrieval practice not only contributed to an increase in student performance in AO1 questions, but also in AO2 and AO3 which had not been specifically targeted in the intervention. Such finding mirror those reported by Roediger & Butler (2011), who found that the use of retrieval practice improved students ability to transfer knowledge to new contexts.

Student views on the impact of the intervention

To further evaluate their views of the practice, students in the intervention group were asked to complete an online questionnaire, which asked them to reflect on the intervention they had undertaken. Despite typically poor return rates for questionnaires being reported, (Bell, 1999) the majority of students opted to take part in this stage of the research and the response rate to the questionnaire was 80%. All parts of the questionnaire were completed by all students. No student ticked the same value for each response, suggesting that questions were answered thoughtfully. To summarise the quantitative data from my questionnaires, I have used tables and graphical representation to show the percentages of students that gave a particular response to give clarity to the reader. In addition, some examples of student responses have been quoted to illustrate the conclusions I have drawn.

Due to the Covid pandemic, an online questionnaire was used in place of the planned face-to-face group interviews with students to elicit their views on the intervention. All of the students in the cohort are classified as EAL with the range of English language skills from basic to advanced as defined by IELTS descriptors. The College ethos is that all lessons are conducted in English and therefore all students were presented with the questionnaire in only this language. The wording of the questions used was discussed with the Head of EAL at the college in terms of the language content, prior to issuing it, and amendments were made to the wording of questions where appropriate.

Some, but not all students took the opportunity to provide written responses to questions when asked if they would like to add any comments to questions. Due to the limited number and range of comments given, it was not possible to produce a matrix of responses but the comments given have been used to support conclusions made where appropriate.

The intervention was carried out with two Year 2 classes, taught by different teachers. The initial question used in questionnaire asked the students what they understood the purpose of the retrieval practice quizzes, that they had participated in, to have been. All students answered the question and their responses showed that their understanding of the intervention was in line with the purposes of the retrieval practice quizzes as we had used them in college. In explaining the nature and purpose of the weekly tests, both teachers had effectively conveyed the same message to their classes.

Students commented that the quizzes were useful to help them to “keep on track and remember the content”. One student commented that they

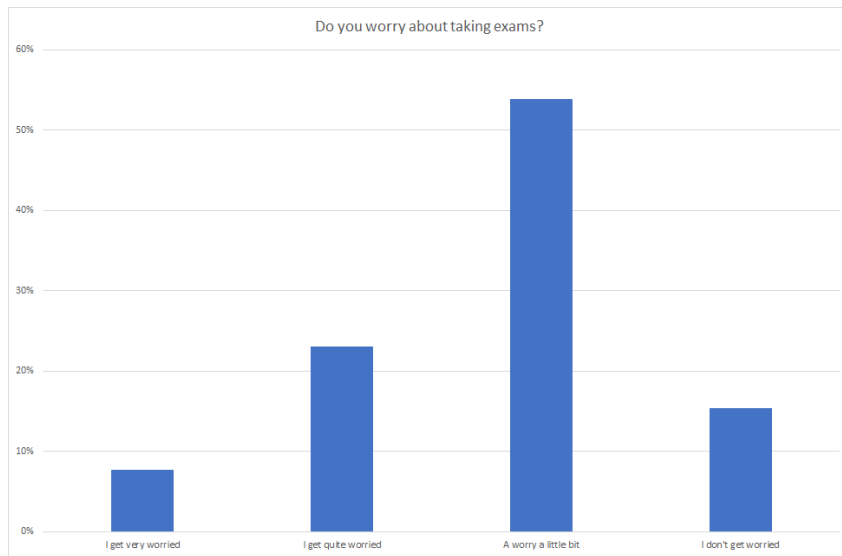
“tend to forget things which we have gone through in our class, so I think that the purpose is to recall in my memory any details and topics which we went through earlier”.

Another student commented that the purpose of the quizzes was to “move the learn[t] material into the long-term memory”, which suggests that using the cognitive theory memory model, as discussed in the literature review, when the project was outlined initially, had helped this particular student to understand the purpose of the project.

The impact of additional testing on student anxiety?

With the strong emphasis on the outcomes of high stakes testing in our education system, I was concerned that the introduction of additional tests for my students may increase their test anxiety. The questionnaire was used to gain responses from students regarding how they viewed examinations. Students were asked if they worried about examinations in general and if they worried about the weekly quizzes used in the intervention.

One student reported that they felt very nervous about examinations, and 77% of respondents said that they experienced moderate levels of worry when they sat exams with one student indicating that they got very worried about examinations., figure 11. All of the students said that as a result of the weekly testing, they felt less worried about their chemistry examinations compared to other subjects. Although a range of factors may contribute to student levels of anxiety when approaching a chemistry exam compared to other examinations, I can conclude that the weekly retrieval practice and additional testing did not make them more worried about chemistry examinations than usual. These responses support the findings of Agarwal et al. (2014) that use of retrieval practice *reduces* test anxiety, and that students who participated in frequent retrieval practice exercises, such as quizzes, felt less anxious about upcoming tests.

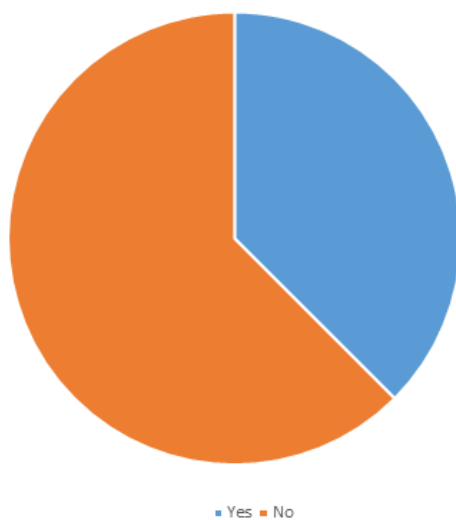


I get very worried	8%
I get quite worried	23%
A worry a little bit	54%
I don't get worried	15%

Figure 11 Student responses to exam anxiety

In terms of the student's attitude to the weekly retrieval quizzes, 63% of the respondents said that they were not worried at all by the tests, with 31% indicating that they were only a little worried by them. This shows that the majority of students were not negatively impacted by the additional weekly quizzes in terms of test anxiety. One student did respond to indicate that they did worry about the weekly tests.

Do you worry about doing the retrieval quizzes?



Yes	37%
No	63%

Figure 12 The impact of additional testing on student anxiety

As the questionnaires were anonymous, this individual cannot be identified, therefore, I do not know if the responses was from a student is who generally worries about any form of assessment or whether they were concerned about the quizzes in particular, despite their low-stakes status. Students were given the opportunity to expand on their answers using a comment box. However, only students who were not worried made comments, highlighting the disadvantage of questionnaires over face to face interviews, as they do not allow for follow up questions.

Of those students who did make additional comments and expand on their selected answer, one student commented that “I like doing those, because I can understand what I should work on”. Another student commented that

“the grade is just for me and it’s just a means of seeing what I can remember and what I can’t do so it doesn’t worry me”.

These positive comments regarding the impact of the quizzes on the students are in line with the purpose of the intervention and suggest that little or no negative impact had been felt by the majority of participants.

Independent use of retrieval practice

I made use of the questionnaire to ask how my students usually prepared for examinations. The aim of this question was to see if using retrieval practice questions in class had influenced how students prepared to examinations an if they were using the techniques independently. Students reported using a range of techniques in preparation for exams such as reading back through their own notes or doing past paper questions, with a small number (12%) reporting that they did not really prepare for examinations.

Although not part of the intervention described here, In collaboration with my teaching colleague, a set of retrieval questions had been developed, following the outcomes of this project and had been provided to the students who had been in the intervention group to support students preparation for internal examinations. The questions were formulated in the same manner as the weekly quizzes used in the intervention strand of this project, and were designed to be done daily and independently by the students. Therefore the final question on the student questionnaire was included to elicit how students were using retrieval practice independently post intervention.

Of those students who responded to the questionnaire, 38% of students reported that they did all of the materials set, on the day that they were set with 57% of respondents completing all or most of the exercises independently, figure 13.

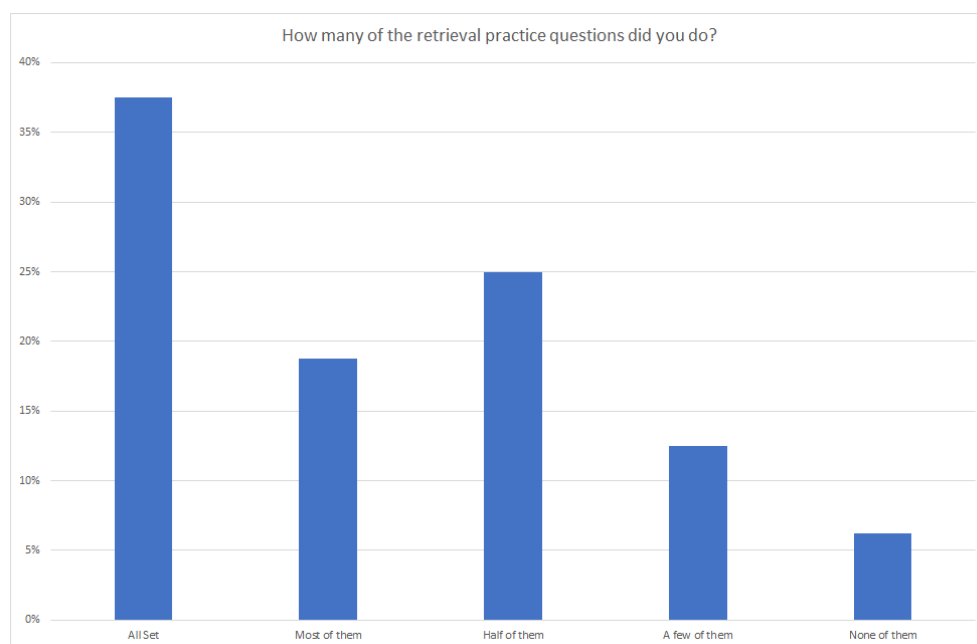


Figure 13 Student independent use of retrieval practice questions.

This data shows a difference to the findings reported by Karpicke et al (2009) that who found that majority of students do not participate in retrieval practice despite knowing its effectiveness. 19% of my students said that they did very few or no retrieval practice questions independently.

At this stage of the project the college was in lockdown so students could not be asked directly what type of examination preparation, if any, that they did instead. This data highlights one of the drawback of questionnaires that was discussed in the methodology, that they do not allow students to elaborate on their answers or allow for follow up questions to probe responses.

Although 56% of students reported that they were using retrieval practice style questions, the majority were not spacing their learning. Only 18% of respondents did questions daily with the rest of the students reported that they either leaving them to build up and did them in batches or reported that they them all in one sitting, cramming their revision, figure 14.

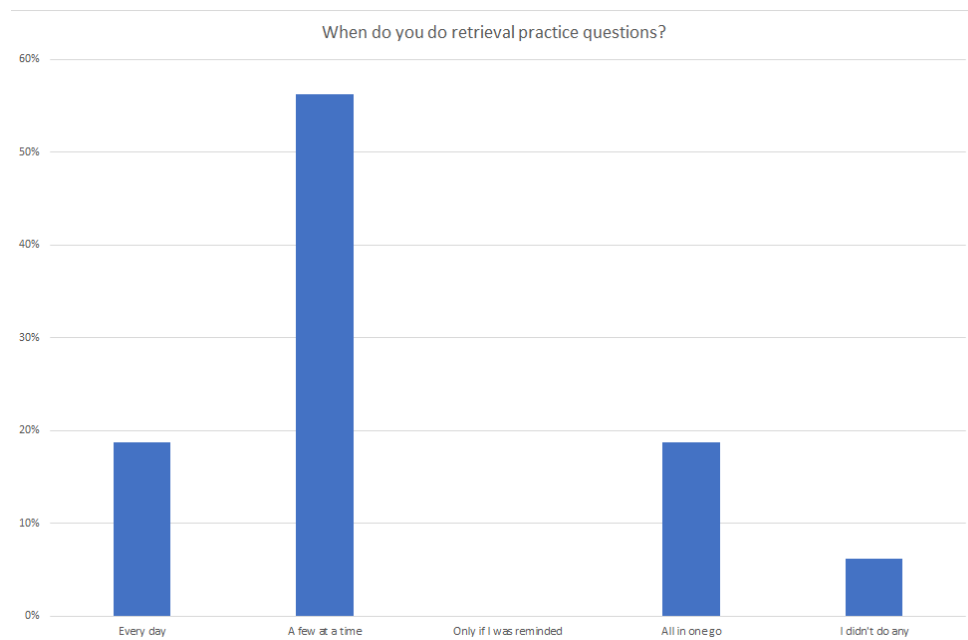


Figure 14 How students space retrieval practice questions

These findings are in line with studies reported by Karpicke et al. (2009) that despite understanding the benefits of using retrieval practice techniques over other methods of study, students do not engage in regular retrieval practices independently.

Conclusion and implications

In my literature review, a number of benefits of retrieval practice were highlighted and the outcomes of this work have been considered in light of these suggested benefits. The conclusions from this work are summarised below before being discussed in more detail.

- At the start of the project the use and understanding of retrieval practice was limited in across the science department in my college.
- That learning of 'basic facts' such as definitions, was considered to be trivial by teachers in the department and that no checks were being made that such learning was taking place. There was a general assumption that it was something that students did independently.
- Following the intervention, student performance improved in AO1, AO2 and AO3 questions. Statistical analysis of student data suggests that the improvement in performance is significant.

When planning this research, I set out to answer two key questions. Firstly, I wished to consider how teachers and students were making use of retrieval practice strategies within the college at the start of the project.

Through the teacher interviews conducted in the college, the teachers were in agreement that they believed there is a significant body of content that needs to be learnt by students in all science subjects. Teachers agreed that learning content is required in order for students to be able to answer exam questions, either directly or through application of the subject knowledge. Prior to the intervention strategy retrieval practice was not routinely carried out in the department. Instead, learning was considered to happen without explicit support or guidance and no formal practices were in place to check that this was indeed the case.

Validity of conclusions

Consideration was given to the validity of the data collected prior to carrying out the interviews, in terms of the conflict between my role as a senior teacher within the college and a participant's view that there was a 'right answer' to the questions. Cannell (1968) reported that validity is a problem when carrying out interviews and their findings were taken into consideration during the design of the data collection process. By using a semi-structured approach to the interviews, I hoped to gain depth and richness to the questions posed, that would not have been possible using a more anonymous questionnaire style of data collection.

The studies of Scheurich (1995), were also considered when conducting the interviews, in terms of aiming to reduce the influence of myself as the interviewer on the responses given by the interviewee. Whilst the barrier of my role as a line manager could not be eliminated, the effect of this on the perception that there was a 'right answer' was addressed by meeting colleagues informally over a coffee. In addition, the initial question aimed to put the participant at ease, by asking them to talk about themselves and their role in the college with further questions about their understanding of retrieval practice and their current methods employed then introduced.

Due to the time constraints of this dissertation, I did not triangulate the data collected from an individual participant, by employing more than one data collection method per person. Whilst it is appreciated that the sample size is very small, the individuals were interviewed separately and did not share their experiences with each other during the process and therefore their responses were their own and can be compared with one another.

It has been argued that due to the interpersonal nature of face to face interviews, that the interviewer influences the data Hitchcock and Hughs (1989). Whilst this effect could not be eliminated in the nature of the method in which the data was collected, it was taken into consideration before the interview in an attempt to minimise its effect.

In addition, an attempt was made to control the reliability of the data by using the same format and sequence of words and questions for each interview conducted as suggested by Silverman (1993). It is again appreciated that this would not guarantee to control the interview, but Oppenheim (1992, p 147) argued that the wording used is particularly important when gaining participants perceptions and attitudes, rather than factual responses. He also argued that changing the wording undermines the reliability of data collected as the questions are then no longer the same and responses can therefore not be compared.

What was the impact of an intervention on retrieval practice on student performance outcomes in AO1 questions?

The second question considered throughout this project was the impact of the intervention using retrieval practice on student performance in AO1 questions. A retrieval practice intervention was designed and carried out weekly to review work that had been covered prior to that week. Students knew that they were low stakes and feedback from student questionnaires following the intervention showed that they did not prepare for them. The quizzes were quick to formulate using search parameters for AO1 questions with tools such as Examwizard (Edexcel).

Data analysed in this project comparing student performances in AO1, AO2 and AO3 questions pre and post intervention showed that the use of retrieval practice appears to improve performance in AO1 questions. These findings are in line with those reported by Rowland (2014).

Data also showed that performance in AO2 and AO3 questions increased, despite these assessment objectives not being targeted in the intervention strategy. I have found a bigger impact following our emphasis on knowledge retrieval on our AO2 rather than our AO1 student outcomes. These findings suggest that retrieval practice also supported the student's ability to transfer of knowledge.

These findings have significance in my department for both teachers and students in terms of enhancing student performance in AO2 and AO3 questions which test their ability to make links between new information and existing knowledge. Roher et al (2010) and Butler (2010) have reported similar findings.

Analysis of student data following the intervention support the studies carried out by cognitive psychologists , such as Bjork (1999), that testing enhances learning. Attempting to retrieve information from memory leads to a host of direct and indirect benefits with respect to the retention, understanding, and application of knowledge and skills.

Through the intervention developed in this project based on retrieval practice strategies, I aimed to show students what they know (what they are able to recall from memory) and also what they aren't able to recall, that that they could prioritise the next steps in their learning. From the body of literature supporting the use of low-stakes quizzes for this intervention I did not make formal records of the student outcomes of the weekly tasks. However, through the design of the use of the quizzes freed up to circulate the room and but I was able to identify patterns when looking at students books to inform future planning.

Responses from the student questionnaires indicted that although students were positive about and recognised the benefits of retrieval practice, but in line with the findings of Roediger (2011), students do not use this strategy themselves independently. Despite consistent findings across research into memory that greater effects result from testing than from restudying (Einstein et al, 2012, Karpicke and Roediger, 2008).

Validity of conclusions

Consideration has been given to the small cohort size when drawing conclusions in this work and the limits of my ability to make claims about the findings. Due to the small sample size it is appreciated that individual responses, particularly outliers, have the potential to have a large impact on the overall outcome. Two separate, but identical questionnaires were sent out to the two different intervention groups, to enable comparisons between the groups to be made. Analysis of the responses and comments made from each group demonstrated consistency between the groups in their understanding of the purpose of the intervention. The responses also highlighted the strength of the collaboration between the teachers in conveying the same messages to the different teaching groups.

Student views of the intervention

Testing has become a bad word in our society, largely because our educational system places a strong emphasis on high-stakes assessments and standardized evaluations. For many students in my college — perhaps even for *most* students — I have observed that getting high test scores is their primary goal. One of my concerns regarding the use of retrieval practice as a strategy to improve student performance was the impact that increasing the amount of testing may have on student anxiety.

Responses to the student questionnaires indicated that students in the college viewed the tests used in the intervention as learning tools rather than as assessment tools with 100% reported to feeling less anxious about chemistry tests compared to other subjects. An advantage of testing as a learning tool from this study, is the reported reduction in test anxiety by students in the intervention group.

The outcomes and findings of this work have highlighted to me that my role as a teacher, is not to focus teaching on transferring information to students' long-term memory, but also to support students to be able to retrieve that information when required. In contrast to my initial teacher training and INSET throughout my career, where the focus has been on how to effectively get information across to students, this project has highlighted that focus should also be given to supporting students in retrieving information. In agreement with Agarwal (2019) I shouldn't just focus on getting information into students' mind but instead ask, '*how can we get that information out of their mind?*'

This work has highlighted that within my classroom, and across my department we could be taking greater advantage of testing as a *learning tool*. Rather than using testing to measure what students have been learned, testing has the potential to improve learning itself. Therefore, within the college our aim will be to start to view and use tests, not as a means to an end, but as a powerful way to equip our students with the knowledge and skills we want them to possess.

The student sample taken do not reflect the entire population of students taking A' levels nationally. However, analysis of the similarities between the control group and the intervention group indicate that they are representative of my college as a whole. Therefore whilst conclusions cannot be drawn as to the impact of this study on all A' level students, generalisations can be drawn with regards to the impact of the study on students within my college.

When drawing conclusions from my data I have considered the limitations that affect my study. This intervention was conducted in only one college with one cohort of students, and therefore no generalisations can be made from my findings. The intervention shows an improvement in student performance outcomes in AO1, AO2 and AO3 questions.

The use of triangulation was intended go some way towards decreasing the disadvantages associated with my research methods; strengths and weaknesses of my research methods are discussed in more detail in the methodology section. It is important to bear in mind that the students were aware that an intervention was taking place; this knowledge may have affected the validity of my data. In spite of these considerations, the process of evaluating my teaching practice has been of great value to me and this study has had lasting effects within the wider context of my teaching.

Unplanned outcomes

I hope that what we have put in place will benefit our students in a number of ways including:

- Greater ability to retrieve and use information from previous topics
- Deeper understanding of subject content
- Greater ability to identify links between topics
- Timely feedback to have a greater impact on learning
- Regular opportunity for improvement
- Improved exam results

In addition to the impact of retrieval practice on student performance outcomes in AO1 questions, data analysis indicates that improvement has also been made in AO2 and AO3 questions, despite these not being the focus of the intervention.

Teachers reported that they gained much from working collaboratively. There is little doubt that without working with other teachers the range of ideas that were offered throughout the analysis and intervention planning would have been narrower in scope.

The identification of AO1 questions to develop the weekly quizzes has assisted me in making my own subject specific pedagogical knowledge more fluent. In lesson planning meetings with my fellow chemistry teacher, we have commented on how our planning focus has changed since the start of the intervention, and we are making the learning outcomes more explicit.

The next steps

Although analysis of the development of metacognition by students through the use of retrieval practice is beyond the scope of this work at the stage, Roediger & Butler (2011) have reported that retrieval practice may be a tool to enable students to identify gaps in their knowledge.

The design of the feedback stage of the intervention was to provide the correct answer for students to self-correct their work. I wanted the students to be aware if they had retrieved the wrong information, so that they could correct it and the intervention was designed so that feedback would be given immediately after students had actively tried to recall the information independently. As students become more aware of their learning, an additional benefit of retrieval practice may be to support the development of students' metacognitive skills. Agarwal (2017) has stated that the use of feedback in retrieval practice "isn't just so students know if they got their answer correct or incorrect, it adjusts students' metacognition", helping students to become "better at estimating or judging what they know and what they don't know." Despite a large body of evidence suggesting that the use of metacognitive strategies may have a significant effect on a student's progress (for example Petty et al 2019), studies linking retrieval practice to the development of metacognition skills have not been reported within the literature at the current time of writing. The aims within my department over the next academic year is to develop our students use of self-monitoring and so help them develop an awareness of learning and how it relates to the behaviours, methods, and techniques that they use.

References

Adescope, O.O., Trevisan, D.A., and Sundararajan, N. (2017). Rethinking the use of tests. A meta-analysis of practice testing. *Review of Educational Research*. 87(3), 659-701.

Agarwal, P. K. (2017) in <https://www.cultofpedagogy.com/retrieval-practice/> [Accessed on 8/6/2020].

Agarwal, P. K. (2019). Retrieval Practice & Bloom's Taxonomy: Do Students Need Fact Knowledge Before Higher Order Learning? *Journal of Educational Psychology*, 111(2), pp. 189–209.

Agarwal, P. K., D'Antonio, L., Roediger, H. L., McDermott, K. B., and McDaniel, M. A. (2014). Classroom-based programs of retrieval practice reduce middle school and high school students' test anxiety. *Journal of Applied Research in Memory and Cognition*, 3, 131–139.

Anderson, R. (1977). 'The notion of schemata and the educational enterprise: General discussion of the conference', in R Anderson, R Spiro & W Montague (eds), *Schooling and the acquisition of knowledge*, Erlbaum, Hillsdale N.J., p. 415-431.

Angus, S. D. & Watson, J. (2009) Does regular online testing enhance student learning in the numerical sciences? Robust evidence from a large data set. *British Journal of Educational Technology* vol:40 iss:2 p:255 -272

AQA, (2015) AS and A Level Chemistry Specification, [Online]. Available from: <https://filestore.aqa.org.uk/resources/chemistry/specifications/AQA-7404-7405-SP-2015.PDF>

Atkinson, R and Shiffrin, R. (1968), 'Human memory: A proposed system and its control processes', in K Spence (ed), *The psychology of learning and motivation: Advances in research and theory*, vol. 2, Academic Press, New York, pp. 89-195.

Baddeley, A 1983, 'Working memory', *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, vol. 302, no. 1110, pp. 311-324.

Baghdady, M., Carnahan, H., Lam, E., & Woods, N. (2014). Test-enhanced learning and its effect on comprehension and diagnostic accuracy. *Medical Education*, 48(2), 181–188

Barnett, S., & Ceci, S. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin* 128, 612-37.

Barrett, L, Tudage, M and Engel, R. (2004). 'Individual differences in working memory capacity and dual process theories of the mind', *Psychological Bulletin*, vol. 130, no. 4, pp. 553-573.

Barry, C. L., Horst, S. J., Finney, S. J., Brown, A. R., & Kopp, J. P. (2010). Do examinees have similar test-taking effort? A high-stakes question for low-stakes testing. *International Journal of Testing*, 10(4), 342-363.

Bell, J. (1999), *Doing your research project. A guide for first time researchers in education and social science*. 3rd Edition. OUP, p112

BERA (2011). *Ethical Guidelines for Educational Research*. Ethical Guidelines for Educational Research.

<http://www.bera.ac.uk/researchers-resources/publications/ethical-guidelines-foreducational-research-2011>

[Accessed 14/3/2020]

Binks, S. (2018), *Journal of Professional Nursing* 34, 205–210

Bjork, R. (1994). Institutional impediments to effective training. In D. Druckman, & R. A. Bjork (Eds.), *Learning, remembering, believing: Enhancing individual and team performance* (pp. 295–306). Washington, DC: National Academy Press.

Bjork, R. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher, & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 435–459). Cambridge, MA: MIT Press.

Bjork, R. A., & Bjork, E. L. (1992) A new theory of disuse and and old terhory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shrifin (Eds.), *From Learning Processes to Cognitive "Processes: Essays in Honor of William K. Estes, Vol 2*, p 35-67. Hillsdae, NJ: Erlbaum.

Black, P. J. And Wiliam, D. (1998). Assessment and classroom learning', *Assessment in Education: Principles, Policy and Practice*, 5 1: 7-74

Bliss, J., Monk, M. and Ogborn, J. (1983) *Qualitative Data Analysis for Educational Research*. London: Croom Helm.

Borg, W. R., & Gall, M. (1979). *Educational Research An Introduction* (3rd ed.). New York Longman.

Braun, V., Clarke, V., & Gray, D. (2017). *Collecting qualitative data: a practical guide to textual, media and virtual techniques*. Cambridge University Press

Brookhart, S. M. (2008). *How to give effective feedback to your students*. Alexandria, VA: Association for Supervision and Curriculum Development.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How People Learn: Brain, Mind, Experience, and School*. Washington, DC: National Academy Press.

Brünken, R., Plass, J & Leutner, D. (2003). 'Direct measurement of cognitive load in multimedia learning', *Educational Psychologist*, vol. 38, no. 1, pp. 53-61.

Butler, A. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1118–1133.

Butler, A. C., Godbole, N., & Marsh, E. J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology*, 105(2), 290–298.

Butler, A., & Roediger, H. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, 19(4/5), 514–527.

Butler, A., & Roediger, H. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, 36(3), 604–616.

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology*, 19, 619-636.

Cannell, C. F., (1968). *The influence of interviewer and respondent; psychological and behavioral variables on the reporting in household interviews*. APA 6th Edition.

Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U.S. history facts. *Applied Cognitive Psychology*, 23, 760–771.

- Chan, J.C.K. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language*, 61(2), 153-170
- Clark, R, Kirschner, P and Sweller, J. (2012). 'Putting students on the path to learning: The case for fully guided instruction', *American Educator*, Spring, pp. 6-11.
- Cochran Smith, M., & Lytle, S. S. (2009). *Inquiry as stance: Practitioner research for the next generation*. New York: Teachers College Press.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education*. (6th, Ed.) Education (Vol. 55,). Routledge.
- Cole, J. S., & Osterlind, S. J. (2008). Investigating differences between low and high-stakes test performance on a general education exam. *Journal of General Education*, 57(2), 119-130
- Cooper, G & Sweller, J. (1987). 'Effects of schema acquisition and rule automation on mathematical problem-solving transfer', *Journal of Educational Psychology*, vol. 79, no. 4, pp. 347-362.
- Corder, G., & Foreman, D. (2009) *Nonparametric Statistics for Non-Statisticians*. Hoboken: Wiley.
- Cowan, N. (2001). 'The magical number 4 in short-term memory: A reconsideration of mental storage capacity', *Behavioural and Brain Sciences*, vol. 24, no. 1, pp. 87-114.
- Daniel, D. & Broida, J, (2004). Using Web-Based Quizzing to Improve Exam Performance: Lessons Learned. *Teaching of Psychology*, 31 p 207-208
- Day, S. B., & Goldstone, R. L. (2012). The import of knowledge export: Connecting findings and theories of transfer of learning. *Educational Psychologist*, 47(3), 153-176.
- De Houwer, J., Barnes-Holmes, D. and Moors, A. (2013). What is learning? On the nature and merits of a functional definition of learning, *Psychon Bull Rev.* [https://ppw.kuleuven.be/okp/_pdf/DeHouwer2013WILLOT.pdf. Retrieved June 8, 2019].

Denscombe, M. (2007). *The good research guide: For small-scale social research*. Buckingham Open University Press.

De Jong, T. (2010). 'Cognitive load theory, educational research, and instructional design: Some food for thought', *Instructional Science*, vol. 38, no. 2, pp. 105-134.

DeLeeuw, K & Mayer, R. (2008). 'A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load', *Journal of Educational Psychology*, vol. 100, pp. 223-234.

Dobson, J., & Linderholm, T. (2015). Self-testing promotes superior retention of anatomy and physiology information. *Advances in Health Sciences Education*, 20(1), 149–161.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4-58.

Einstein, G., Mullet, H., & Harrison, T. (2012). The testing effect: Illustrating a fundamental concept and changing study strategies. *Teaching of Psychology*, 39(3), 190–193.

EFF (2018) <https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/feedback/> [Accessed on 24/10/2019].

Eklöf, H. (2006). Development and validation of scores from an instrument measuring student test-taking motivation. *Educational and Psychological Measurement*, 66(4) 643-656.

Eklöf, H. (2010). Skill and will: test-taking motivation and assessment quality. *Assessment in Education: Principles, Policy and Practice*, 17(4), 345-356.

Eklöf, H., & Nyroos, M. (2013). Pupil perceptions of national tests in science: Perceived importance, invested effort, and test anxiety. *European Journal of Psychology of Education*, 28(2), 497-510.

Farrimond, H. (2017) *The ethics of research*, The Bera/SAGE Handbook of Educational Research, Eds: Dominic Wyse, Emma Smith, Larry E. Suter and Neil Selwyn. SAGE Publications: Thousand Oaks, CA.

Fox, M., Martin, P., & Green, G. (2007). *Doing Practitioner Research*. Wiltshire: SAGE Publications.

Gerjets, P., Scheiter, K & Cierniak, G. (2009). 'The scientific value of cognitive load theory: A research agenda based on the structuralist view of theories', *Educational Psychology Review*, vol. 21, no. 1, pp. 43-54.

Glaser, R., & Chi, M. T. (1988). Overview. In *The Nature of Expertise* (pp. xv-xxvii). Hillsdale: Erlbaum

Harlen W, Deakin Crick R (2002). A systematic review of the impact of summative assessment and tests on students' motivation for learning.

http://eppi.ioe.ac.uk/cms/Portals/0/PDF%20reviews%20and%20summaries/ass_rv2.pdf?ver=2006-03-02-124715-153 [Accessed on 24/02/2020].

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81-112

Hattie, J. (2008). *Visible Learning: A synthesis of over 800 meta-analyses relating to achievement*, New York, Routledge

Hitchcock, G., & Hughes, D. (1989). *Research and the teacher: A qualitative introduction to school-based research*. London and New York: Routledge.

Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning & Verbal Behavior*, 10, 562-567.

Jacoby, L., Walhheim, C., & Coane, J. (2010). Test-enhanced learning of natural concepts: Effects on recognition memory, classification, and metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(6), 1441–1451.

Jaeger, A., Eisenkraemer, R., & Stein, L. (2015). Test-enhanced learning in third-grade children. *Educational Psychology*, 35(4), 513–521.

Johnston, J. (2012). Planning for research. In J. Oversby (Ed.), *The ASE Guide to Research in Science Education* (1st ed., pp. 158–178). Hatfield, Herts: The Association for Science Education.

Jones, H.E. (2019). The effects of examination on the performance of learning. *Arch. Psychol.* 10, 1–70

Kang, S., Lindsey, R., Mozer, M., & Pashler, H. (2014). Retrieval practice over the long term: Should spacing be expanding or equal-interval? *Psychonomic Bulletin & Review*, 21(6), 1544–1550.

Kang, S., McDermott, K., & Roediger, H. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, 19(4/5), 528–558.

Karpicke, J. D. (2009). Metacognitive control and strategy selection: deciding to practice retrieval during learning. *Journal of Experimental Psychology: General*, 138(4), 469.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331(6018), 772–775.

Karpicke, J. D., Butler, A. C., & Roediger, H. L. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory*, 17(4), 471-479.

Karpicke, J., & Roediger, H. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966–968.

Kelly, L. (2002). What is learning ... and why do museums need to do something about it? A paper presented at the *Why Learning? Seminar*, Australian Museum/University of Technology Sydney, 22 November. <https://australianmuseum.net.au/uploads/documents/9293/what%20is%20learning.pdf>. [Accessed on 17/06/2020].

Kibble, J. D. (2017). Best practices in summative assessment. *Advances in Physiology Education*, 41(1), 110–119.

Knekta, E. (2017). Are all pupils equally motivated to do their best on all tests? Differences in reported test-taking motivation within and between tests with different stakes. *Scandinavian Journal of Educational Research*, 61(1), 95-111.

Kolb, D. A., (1984) *Experiential Learning: experience as the source of learning and development*. Englewood Cliffs, Prentice Hall.

Kornell, N., & Bjork, R. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin and Review*, 14(2), 219–224.

Koshy, V. (2010). *Action Research for Improving Educational Practice* (2nd ed., p. 157). London: Sage Publications Ltd.

Laberge, D. and Samuels, J. (1974). 'Toward a theory of automatic information processing in reading', *Cognitive Psychology*, vol. 6, no. 2, pp. 293-323.

Larsen, D.P. et al. (2009). Repeated testing improves long-term retention relative to repeated study: a randomized, controlled trial. *Med. Educ.* 43, 1174–1181

Leahy, S, Lyon, C, Thompson, M & Wiliam, D. (2005). Classroom assessment: minute by minute, day by day, *Educational Assessment*, 63(3), 19–24

Leppink, J, Paas, F, van Gog, T, van der Vleuten, C & van Merriënboer, J 2014, 'Effects of pairs of problems and examples on task performance and different types of cognitive load', *Learning and Instruction*, vol. 30, pp. 32-42.

Liu, O.L., Bridgeman, B., & Adler, R.M. (2012). Measuring learning outcomes in higher education: motivation matters. *Educational Researcher*, 41(9), 352-362.

Locke, EA & Latham, G. P. (1992). *A Theory of Goal Setting and Task Performance*, Englewood Cliffs, NJ, Prentice Hall

Lovell, O. (2018). *Sweller's Cognitive Load Theory in Action* (In Action Series) Published by John Catt

Marsh, G. (2005). *Illustrating Primary Science: A teacher's use of representations*. University of London.

Marvasti, A. & Freie, C. (2017) in *The BEAR/SAGE Handbook of Educational Research*, Wyse, D., Selwyn, N., Smith, E. & Suter, L. E. SAGE Publications Ltd

Mayer, R, Hegarty, M, Mayer, S and Campbell, J. (2005). 'When static media promote active learning: Annotated illustrations versus narrated animations in multimedia instruction', *Journal of Experimental Psychology: Applied*, vol. 11, no. 4, pp. 256-265.

McDaniel, M., Anderson, J., Derbish, M., & Morissette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, 19(4/5), 494–513.

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L., III. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology*, 103, 399–414.

McDaniel, M., Thomas, R., Agarwal, P., McDermott, K., & Roediger, H. (2013). Quizzing in middle-school science: Successful transfer performance on classroom exams. *Applied Cognitive Psychology*, 27(3), 36–372.

Melton, A. W. (1963) Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*. (2) p 1-21

Menter, I., Elliot, D., Hulme, M., Lewin, J., & Lowden, K. (2011). *A guide to practitioner research in education*. Bodmin: SAGE Publications.

Minstrell, J. (2001). The role of the teacher in making sense of classroom experiences and effecting better learning. In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 121–150). Mahwah, NJ: Erlbaum.

Moreno, R. (2010). 'Cognitive load theory: More food for thought', *Instructional Science*, vol. 38, no. 2, pp. 135-141.

Oppenheim, A. N. (1992) *Questionnaire Design, Interviewing and Attitude Measurement*. London, Pinter.

Paas, F, Renkl, A & Sweller, J. (2003). 'Cognitive load theory and instructional design: Recent developments', *Educational Psychologist*, vol. 38, no. 1, pp. 1-4.

Paas, F, Tuovinen, J, Tabbers, H & van Gerven, P. (2003). 'Cognitive load measurement as a means to advance cognitive load theory', *Educational Psychologist*, vol. 38, no. 1, pp. 63-71.

Pardjono, P., *Active Learning: The Dewey, Piaget, Vygotsky, and Constructivist Theory Perspectives*, *Jurnal Ilmu Pendidikan (Jakarta)*, (2016), Vol.9(3)

Patton, D., (1980) *Qualitative Research & Evaluation Methods Integrating Theory and Practice* 4TH Edition, Sage Publishing

Pellegrino, J. W., & Hilton, M. L. (2012). *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*. Washington, DC: National Academies Press.

Peterson, L and Peterson, M. (1959). 'Short-term retention of individual verbal items', *Journal of Experimental Psychology*, vol. 58, no. 3, pp. 193-198.

Petty, G. (2006). *Evidence Based Teaching*. Cheltenham: Nelson Thornes .

Petty, J., Lundie, D & Goldie, G. (2019) Metacognition in schools: what does the literature suggest about the effectiveness of teaching metacognition in schools? *Journal Educational Review*, 71, p

Piaget, J. (1964). Part I: Cognitive development in children: Piaget. *Development and learning. Journal of Research in Science Teaching*, 2(3), 176–186. <http://doi.org/10.1002/tea.3660020306>

Poortvliet, P. M., Janssen, O., Van Yperen, N. W., & Van de Vliert, E. (2007). Achievement goals and interpersonal behavior: How mastery and performance goals shape information exchange. *Personality and Social Psychology Bulletin*, 33, 1435–1447.

Pyc, M., & Rawson, K. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, 60(4), 437–447.

Rawson, K., & Dunlosky, J. (2012). When is practice testing most effective for improving the durability and efficiency of student learning? *Educational Psychology Review*, 24(3), 419–435.

Reif, F. (2010). *Applying Cognitive Science to Education: Thinking and Learning in Scientific and Other Complex Domains* (A Bradford Book)

Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(1), 233–239

Roediger, H. L. and Butler (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, January 2011, Vol. 15, No. 1, p 20-27

Roediger, H., & Karpicke, J. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1(3), 181–210.

Roediger, H. L., and Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249–255.

Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(1), 233–239.

Rowland, C. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432–1463.

Säljö, R. (1979). 'Learning in the learner's perspective. I. Some common-sense conceptions', *Reports from the Institute of Education, University of Gothenburg*, 76.

Scheurich, J. J. (1995). A postmodernist critique of research interviewing. *International Journal of Qualitative Studies in Education*, 8, 239-252

Schnotz, W & Kürschner, C. (2007). 'A reconsideration of cognitive load theory', *Educational Psychology Review*, vol. 19, no. 4, pp. 469-508.

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 179-187.

Silverman, D. (1993). *Interpreting qualitative data: methods for analysing talk, text and interaction*. London SAGE.

Simon, D. A. & Bjork, R. A. (2001). Metacognition in motor learning, *Journal Experimental Psychology, Learning, memory and cognition*. 27(4), p. 907-912

Stull, A and Mayer, R. (2007). 'Learning by doing versus learning by viewing: Three experimental comparisons of learner generated versus author-provided graphic organisers', *Journal of Educational Psychology*, vol. 99, no. 4, pp. 808-820.

Sundre, D.L. (1999). Does examinee motivation moderate the relationship between test consequence and test performance? Paper presented at AERA meeting, Montreal, Canada, April, 1999. Retrieved from <https://files.eric.ed.gov/fulltext/ED432588.pdf>. [Accessed on 24/10/2019].

Sundre, D.L., & Kitsantas, A. (2004). An exploration of the psychology of the examinee: Can examinee self-regulation and test-taking motivation predict consequential and non-consequential test-performance? *Contemporary Educational Psychology*, 29, 6-26.

Sutton Trust (2020). Social mobility and Covid19. Implications of the Covid-19 crisis for educational inequality. [Accessed at <https://www.suttontrust.com/wp-content/uploads/2020/04/COVID-19-and-Social-Mobility-1.pdf> 26/05/2020]

Sweller, J & Cooper, G. (1985). 'The use of worked examples as a substitute for problem solving in learning algebra', *Cognition and Instruction*, vol. 2, no. 1, pp. 59-89.

Sweller, J, van Merriënboer, J. and Paas, F. (1998). 'Cognitive architecture and instructional design', *Educational Psychology Review*, vol. 10, no. 3, pp. 251-296.

Thelk, A.D., Sundre, D.L., Horst, S.J., & Finney, S.J. (2009). Motivation matters: Using the student opinion scale to make valid inferences about student performance. *The Journal of General Education*, 58(3), 129-151.

Thomas, G. (2009). *How to do your Research Project`* (1st ed., p. 249). London: Sage Publications Ltd.

Tuckman, B., & Harper, B. (2012). *Conducting Educational Research*. Rowman & Littlefield Publishers.

Tulving, E. (1972). 'Episodic and semantic memory', in E Tulving and W Donaldson (eds), *Organization of memory*, Academic Press, New York, pp. 381-403.

van Gog, T. & Sweller, J. (2015) Not New, but Nearly Forgotten: the Testing Effect Decreases or even Disappears as the Complexity of Learning Materials Increases *Educ Psychol Rev* 27, p 247–264

Wellington. J. (2000). *Educational Research. Contemporary Issues and Practical Approaches*. Continuum. London and New York

Wiliam, D. (2009). An integrative summary of the research literature and implications for a new theory of formative assessment. In H Andrade & GJ Cizek (eds), *Handbook of Formative Assessment*, London, Taylor and Francis, 18-40

Wiliam, D. (2010). *Embedded Formative Assessment*, Bloomington, Indiana, Solution Tree.

Wiliam, D. (2017) <https://gregashman.wordpress.com/2017/01/27/cognitive-load-theory-the-single-most-important-theory-for-teachers-to-know/> [Accessed on 19/06/2019].

Wilkinson, D., & Birmingham, P. (2003) *Using Research Instruments: A Guide for Researchers*. Psychology Press

Willingham, D. (2009) *Why Don't Students Like School?: A Cognitive Scientist Answers Questions About How the Mind Works and What It Means for the Classroom*, Published by Jossey-Bass

Wilson, N., & McLean, S. (1994). *Questionnaire Design A Practical Introduction*. Newtown Abbey University of Ulster Press.

Winn, W., & Snyder, D. (1996). Cognitive perspectives in Psychology. In *Handbook of Research on Educational Communications and Technology* (1st ed., pp. 112–142). New York: Simon and Scuster Macmillan.

Wolf, L. F., Smith, J. K., & Birnbaum, M. E. (1995). Consequence of performance, test, motivation, and mentally taxing items. *Applied Measurement in Education*, 8(4), 341–351.

Appendices

Appendix A - Statistical analysis

$H_0: \mu_c = \mu_i$ (c = control group, i = intervention group)

$H_1: \mu_c < \mu_i$ (one-tailed test)

Significance level: $\alpha = 0.05$

Degrees of freedom, = 31

Critical region, $t > 1.697$ ($v = 30$) or Critical region, $t > 1.697$ ($v = 32$)

Values come from the Percentage Points of Student t Distribution table

Under H_0 the test statistic is $t = 4.28$ (AO1)

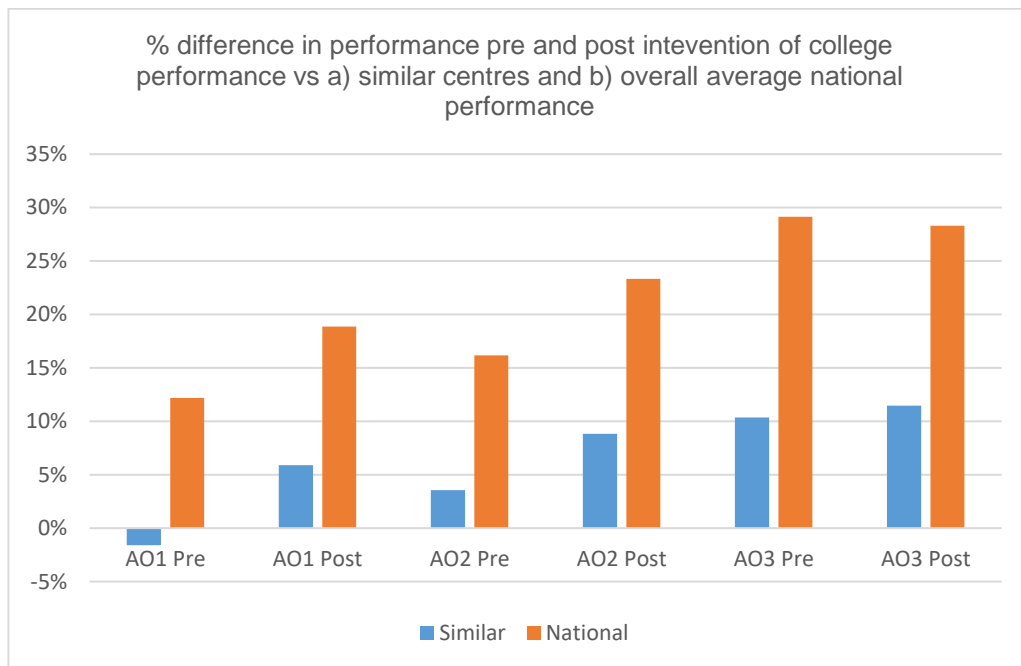
$t = 4.22$ (AO2)

$t = 4.52$ (AO3)

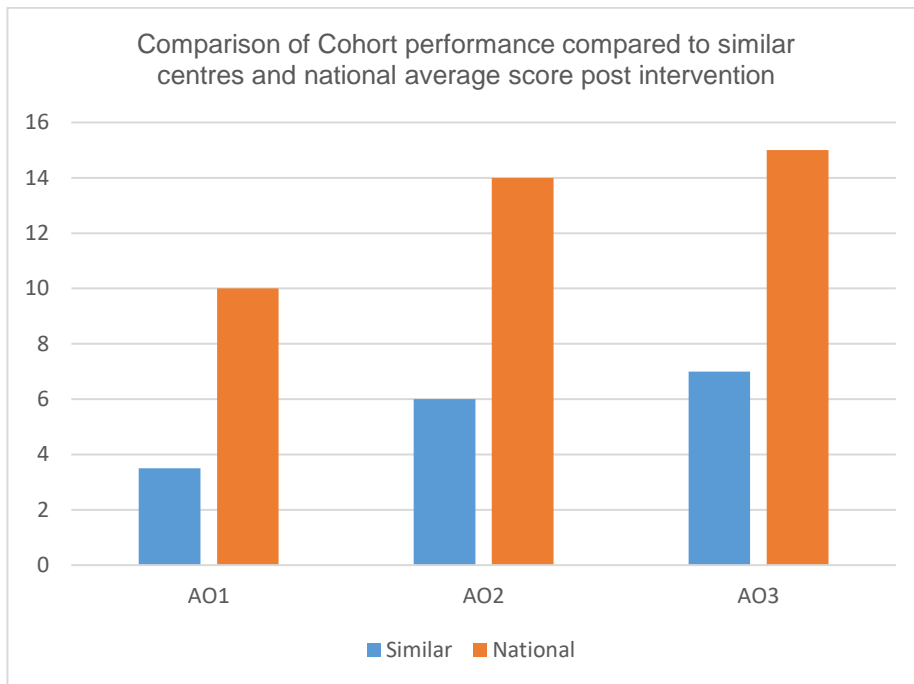
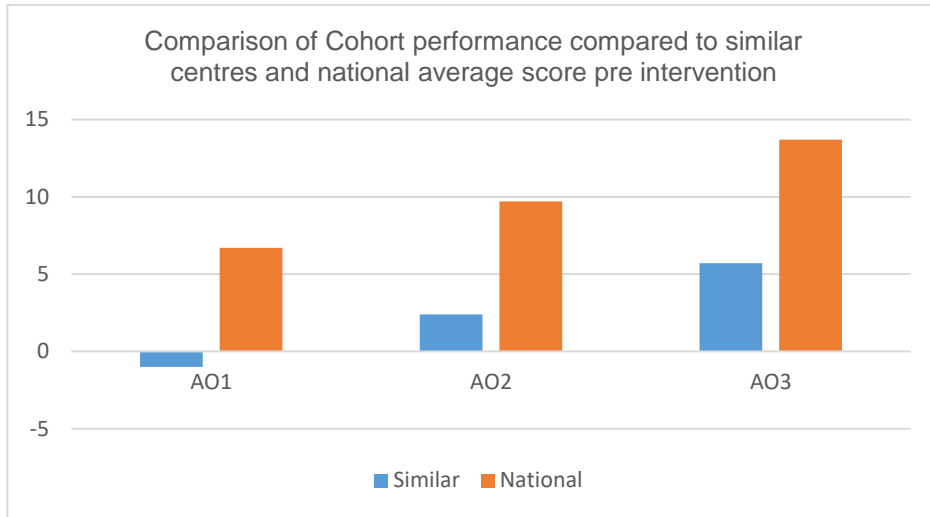
These values all lie in the critical region (for both $v = 30$ or 32) so H_0 is rejected for all three. Thus there is evidence, at the 5% level of significance, to suggest that the intervention group did better than the control group in all three assessment objectives.

Comparison of control group and intervention group compared to national scores for assessment objectives.

The percentage difference in the mean of the control group and the intervention group against a) similar centres (as defined by the exam board) and in also in comparison to b) the national average mean scores for each assessment objective. The data showed that there was an increase in the mean scores post intervention across all AO1, AO2 and AO3 assessment objectives.



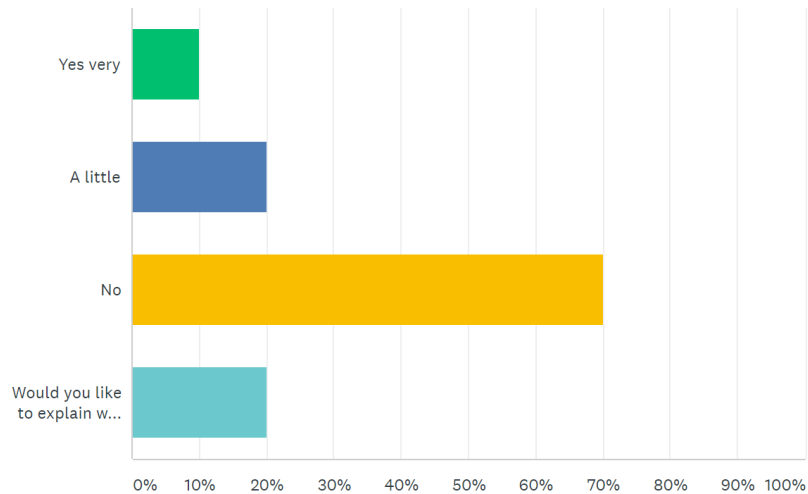
Change in AO1 scores post intervention



Appendix B – Additional Student questionnaire responses to intervention

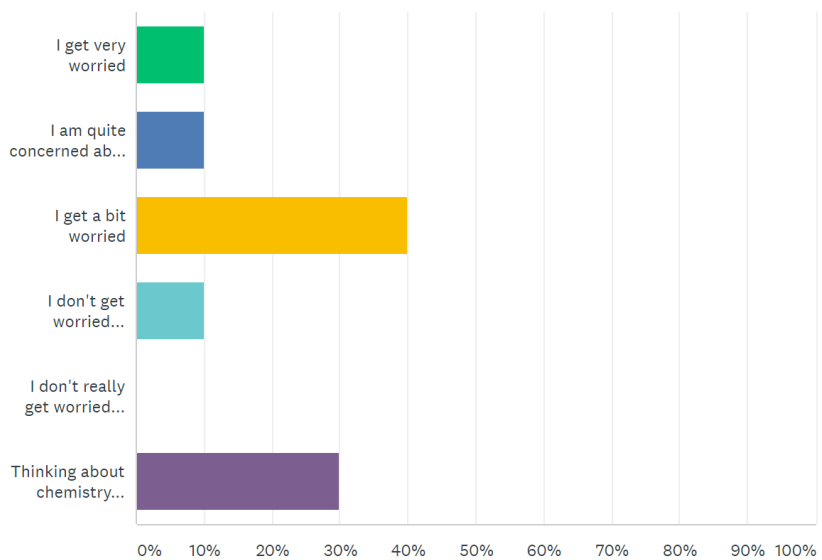
Do you get worried / concerned about doing the weekly retrieval quizzes?

Answered: 10 Skipped: 0



Using this scale - how nervous or worried do you normally get about exams

Answered: 10 Skipped: 0



How do you prepare for exams?

